DOCUMENT RESUME

ED 202 219 FL 012 280

AUTHOR Cervenka, Edward J.; And Others
TITLE Planning Paper 7: Follow-up Studies in Bilingual Education: Issues and Options.
INSTITUTION Abt Associates, Inc., Cambridge, Mass.; Development Associates, Inc., Arlington, Va.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE 1 Sep 80
CONTRACT 400-79-0071
NOTE 112p.

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Bilingual Education; Elementary Secondary Education; *Followup Studies; Measurement Objectives; Measurement Techniques; *Outcomes of Education; Program Evaluation; Program Implementation; Research Design; Research Methodology; Sampling; Statistical Analysis; Student Attrition; Teacher Education

ABSTRACT

Follow-up studies are necessary for investigating the long-term, enduring, anticipated effects of bilingual education. Investigations need to be based on a recognition of their value to educational audiences, the relative merits of research issues, attention to the problem of attrition, proper selection of dependent measures, and the merits of planned follow-up studies. In addressing the need for follow-up studies, this report discusses (1) options or areas of focus for conducting follow-up investigations, (2) follow-up research foci, (3) methodological issues affecting the use of follow-up designs, (4) statistical and analytic issues in conducting follow-up designs, (5) selection of follow-up variables: measurement issues, (6) large scale post hoc follow-up studies, and (7) focused follow-up studies of bilingual education consequences for students. Charts include (1) a typology of follow-up and related designs; (2) potential research areas to investigate dealing with long term consequences of bilingual education on students, long term development of teachers, and long term program implementation and institutionalization; (3) an overall analysis of effectiveness of techniques; and (4) a flow chart of the process by which necessary sample size may be determined. (JK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED202219

PLANNING PAPER 7:
FOLLOW-UP STUDIES IN BILINGUAL EDUCATION: ISSUES AND OPTIONS

Edward J. Cervenka with contributions by
Milton R. Goldsamt and Blair A. Rudes

Prepared under subcontract to Development Associates, Inc. as part of contract #NIE-400-79-0071 awarded to Abt Associates Inc. to conduct a "Study Designed to Assist Planning of Research on Significant Instructional Features in Bilingual Education Programs."

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Development Associates, Inc.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Prepared for:
The National Institute of Education
1200 19th Street, N.W.
Washington, D.C. 20208

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Prepared by:
Development Associates, Inc.
2924 Columbia Pike
Arlington, Virginia 22204

September 1, 1980

FL012280

This report was funded by the National Institute of Education under contract #NIE-400-79-0071, awarded to Abt Associates Inc., Cambridge, MA, and under subcontract held by Development Associates, Inc., Arlington, VA. The contents of this report do not necessarily reflect the views of the Institute or of any other agency of the U.S. Government.

FOREWORD

This report is one of seven submitted by Abt Associates Incorporated (AAI), and Development Associates, Inc. (DA) under subcontract to AAI, to the National Institute of Education, as products of a "Study Designed to Assist Planning of Research on Significant Instructional Features in Bilingual Education Programs", Contract No. NIE-400-79-0071. The reports are intended to assist NIE in its plans for major new research studies in bilingual education. The information provided in Reports 1-5 has been combined with that from other sources by NIE in its construction of a research plan. This plan has been incorporated in two requests for proposals (RFP's) to implement and conduct a major study and three special studies of bilingual instructional features.

The Instructional Features Study was formulated by the Division of Education Part C Coordinating Committee as one of several studies that implement research mandates in the language of ESEA Title VII, Part C. A description of the study (denoted "B-1") is provided in the U.S.H.E.W. Research Plan for Bilingual Education (July, 1979). This planning assistance study was one component of Phase I of the three phase HEW plan.

These reports were prepared as products of Tasks 1-7 of the planning assistance contract. The titles of the reports, and summaries of their contents, are:

1. Working Definitions of Terms for the Bilingual Instructional Features Study, by Sarah Nieves-Squires and Robert L. Goodrich.

   This is a discussion of working definitions of terms for use in the features study. The terms discussed are "bilingual education", "consequences for children", "instructional features", "significant", and "model". Alternative definitions and the implications of each for design are presented. The definitions selected by NIE are intended to guide the research to be conducted.

2. A Bibliography of Significant Instructional Features in Bilingual Education Programs, by Sarah Nieves-Squires, et al.

   This is an annotated bibliography of papers, articles, pamphlets and books that deal with instructional features of bilingual education. The materials

4

are organized by a classification system of features based on a content analysis of the sources surveyed. The report demonstrates that, while many instructional features are discussed, there is little or no empirically based research on their specific consequences for children.

3. Planning Factors for Studies of Bilingual Instructional Features, by Robert L. Goodrich, *et al*.

   This report is based on a review of studies of educational instructional features in both monolingual and bilingual contexts, and on conversations with a large number of researchers and critics. The intent of the report is to summarize the state of the art of bilingual educational features research as a base for designs to be developed by NIE.

4. Tentative Alternative Designs for a Study of Significant Instructional Features in Bilingual Education, by Robert L. Goodrich.

   This report presents alternative study designs and plans for implementation of the instructional features study. It is based on the knowledge base assembled in the three preceding reports. The designs presented are not to be implemented directly by NIE in the RFP(s); rather, they are to be used as one source of information available to the NIE planners, to be factored into the overall design process.

5. Feasibility and Credibility of Bilingual Instructional Features Study Plans: Field Verification, by Sarah Nieves-Squires, Robert L. Goodrich, and Cristina Bodinger-de Uriarte.

   This report summarized the results of 123 open interviews conducted with bilingual practitioners and administrators in five sites: Los Angeles, Miami, New York, Oakland, and Rough Rock, AZ. The questions asked in the survey were designed to elicit responses about the working definitions of terms (Paper 1) and the designs considered (Papers 3, 4). The intent was to test the credibility and acceptability to consumers of alternative study approaches.

6. Issues in Applying Ethnography to Bilingual Classroom Settings, by Blair A. Rudes, Milton R. Goldsamt and Edward J. Cervenka. Prepared under subcontract by Development Associates, Inc.

   This report is based on an extensive review of ethnographic and methodological sources related to bilingual education issues and a conference on current issues in applying ethnographic methods to that domain held at Development Associates on June 24-25, 1980. The report attempted to demonstrate that ethnographic methods have rapidly become established as a new and significant research paradigm for bilingual education research, and for the social sciences generally. Indepth discussions of measurement approaches (including videotape recording), reliability and validity, perceived roles of ethnographers, ways of analyzing ethnographic information, and joint application of ethnographic and quantitative approaches are included in the report.

7. *Follow-Up Studies in Bilingual Education: Issues and Options*, by Edward J. Cervenka, with contributions by Milton R. Goldsamt and Blair A. Rudes. Prepared under subcontract by Development Associates, Inc.

This report examines the issues involved in designing follow-up studies in bilingual education. It reviews conceptual, design, statistical and analytic issues involved in follow-up research, particularly noting the weaknesses and strengths of such designs for bilingual education. Particularly covered are attrition issues, and techniques for identifying and maintaining follow-up samples. The options or areas considered for follow-up research are: the significant bilingual instructional features study, its three related special studies, Part C Studies generally, and certain other selected studies or programs. Large-scale planned and *post hoc* studies are recommended for studying the long term consequences for students, teacher professional development and program implementation and institutionalization. Focused follow-up studies of the consequences of bilingual education for students are strongly recommended.

TABLE OF CONTENTS

Section Page

FOREWORD iii

LIST OF EXHIBITS viii

EXECUTIVE SUMMARY. ix

1.0 INTRODUCTION 1-1

1.1 Purpose of Planning Paper 1-1
1.2 Need for Conducting Follow-up Studies 1-3
1.3 Overview of Follow-up Designs 1-5

2.0 OPTIONS OR AREAS OF FOCUS FOR CONDUCTING FOLLOW-UP INVESTIGATIONS. 2-1

2.1 Significant Instructional Features Study (B-1). 2-1
2.2 Special Studies Associated with B-1 2-2
2.3 Part C Studies. 2-3
2.4 Promising Follow-up Opportunities 2-4

3.0 FOLLOW-UP STUDY RESEARCH FOCI. 3-1

3.1 Long Term Consequences for Students 3-1
3.2 Teacher Long Term Professional Development. 3-3
3.3 Program Implementation and Institutionalization 3-4

4.0 METHODOLOGICAL ISSUES AFFECTING THE USE OF FOLLOW-UP DESIGNS 4-1

4.1 General Considerations. 4-1
4.2 Unanticipated Consequences. 4-3
4-3 Learning Thresholds 4-4
4.4 Use of Comparison Groups. 4-5
4.5 Effects of Time Prior to Follow-up and Thereafter 4-8
4.6 Cross-Sectional Follow-up Designs 4-12
4.7 The Problem of Attrition. 4-15
4.8 Provision for Identifying Follow-up Samples 4-21
4.9 Techniques for Maintaining Follow-up Samples. 4-24
4.10 Related Cost Considerations 4-33
4.11 Baseline Data 4-35
4.12 The Role of Tertiary Data 4-37
4.13 Ethnical Considerations 4-37

TABLE OF CONTENTS (Continued)

Section Page

5.0 STATISTICAL AND ANALYTIC ISSUES IN CONDUCTING FOLLOW-UP DESIGNS. 5-1

5.1 Change Scores 5-1
5.2 Regression Effects. 5-1
5.3 Analysis of Covariance. 5-2
5.4 Size of Effects 5-2
5.5 Trend Analyses and Time Series Analyses 5-3
5.6 Factor or Cluster Analysis. 5-6
5.7 Path Analysis 5-7
5.8 Nonparametric Statistics. 5-8
5.9 Final Thoughts on Statistical Techniques. 5-9

6.0 SELECTION OF FOLLOW-UP VARIABLES: MEASUREMENT ISSUES. 6-1

6.1 Initial Considerations. 6-1
6.2 Consequences for Students 6-2
6.3 Professional Development of Teachers. 6-3
6.4 Program Implementation and Institutionalization 6-4

7.0 LARGE SCALE POST HOC FOLLOW-UP STUDIES 7-1

7.1 Follow-up Research Into Long Term Consequences for Students . . 7-2
7.2 Long Term Professional Development of Bilingual Teachers. . . . 7-4
7.3 Program Implementation and Institutionalization 7-5

8.0 FOCUSED FOLLOW-UP STUDIES OF BILINGUAL EDUCATION CONSEQUENCES FOR STUDENTS 8-1

8.1 Continued Research on Instructional Features. 8-1
8.2 Language Acquisition Follow-up. 8-2
8.3 Language Use Follow-up. 8-3
8.4 Cognitive Development Follow-up 8-4
8.5 Case Study of Personality Factors 8-5
8.6 Case Study Follow-up of Students' Performance in All-English Speaking Classrooms 8-6

9.0 CONCLUSIONS. 9-1

REFERENCES

LIST OF EXHIBITS

Exhibit		Page
1	Typology of Follow-up and Related Designs	1-8
2	Potential Research Areas to Investigate Dealing with the Long Term Consequences of Bilingual Education on Students	3-2
3	Potential Research Areas Dealing with the Long Term Development of Teachers	3-3
4	Potential Research Dealing with Long Term Program Implementation and Institutionalization	3-4
5	Form of the Time-Sequential, Cohort-Sequential, and Cross-Sequential Longitudinal Designs	4-11
6	Summary of Sources of Differences in Three Types of Data When Intact School Groups are Tested	4-13
7	Process by Which Necessary Sample Size May be Determined	4-16
8	Tracing Sources of Information	4-31
9	Overall Analysis of Effectiveness of Techniques	4-31
10	Analysis of Effectiveness of Letters to Individuals in the Potential Sample	4-32

EXECUTIVE SUMMARY

This summary deals with the major points discussed in Planning Paper 7, a presentation of the conceptual, methodological and practical issues and options which exist in planning and designing follow-up studies in the field of bilingual education.

Need for Conducting Follow-up Studies

A long term perspective or interest in the enduring effects of education in the United States could almost be called a national or cultural blind spot. Rare indeed are the studies that follow children through the school system and into college or adult life. In the case of the diverse innovation of bilingual education, a blind spot for long term perspectives on education and human development again appears to exist. Research and policy evaluation continue to insist on the quick, yet perhaps unfair and misleading, demonstration of achievement over a relatively short period of time. The conditions for long-term research are present, especially when one considers that federally funded bilingual education has been around since 1967 through ESEA Title I and III, and with much greater intensity since 1968 with the passage of the Bilingual Education Act.

Longitudinal studies of bilingual education are probably infeasible at the moment, although having the results of well-conceived follow-up studies of existing studies and programs would be a useful interim step. Follow-up studies employing a long term perspective are needed to confirm and further explore largely academic and suggestive accounts of the effects of bilingualism and bilingual education on: language acquisition, cognition, problem solving, educational achievement, personality and character formation, attitudes, career and life choices of individuals. There is a specific need to confirm the existence and consequences of language learning thresholds for dominant and second languages (i.e., L1 and L2). Information is also needed on exit criteria and the subsequent placement of bilingual education students in school system mainstream classes. Furthermore, information is needed on the type of support students require (if any) after they leave bilingual education classrooms. Above all, a need exists to explore further all the meaningful consequences -- anticipated and unanticipated; positive and negative -- of students' participation in types of bilingual education programs.

On another plane, what has been said of long term development of the student, can also appropriately be said for the bilingual teacher. There is a need for a long term perspective on the training and professional development of the teacher. Systematic knowledge on the long term development of bilingual teachers is a significant gap in improving the quality of bilingual teacher education and inservice training.

Finally, the long term perspective, which could be provided by various forms of follow-up study, needs to be applied to programs. Program implementation, and particularly, program institutionalization studies are needed. That is, a long term follow-up will be able to provide a better understanding of how and to what extent programs have been incorporated into local school systems, once external (usually federal) funding has terminated.

This report treats the major descriptive study of Phase II and its accompanying three special studies as the primary current opportunity for follow-up study in bilingual education. However, the authors found it worthwhile to examine the broader area in which follow-up studies in bilingual education may be carried out, viz.:

a. Selected Part C Studies.

b. Other individual studies which offer unusual promise or opportunity.

c. Current new research on students, teachers and programs (i.e., not merely follow-ups of prior studies).

The report argues that there are three primary aspects of current bilingual education which could be suitably examined by means of follow-up research designs: long term consequences for student; teacher long term professional development; and program implementation and institutionalization. Each of these foci represent the principal levels of analysis needed for future follow-up studies. The report offers and elaborates on the scope of illustrative research questions for each of these levels.

This report also supplies a comprehensive review of follow-up research and evaluation methodology, particularly as it may be applied to bilingual education issues. Discussions are provided of alternative longitudinal designs, the need for comparison groups, ways of dealing with attrition, cohort maintenance approaches, and feasible statistical techniques, among other topics.

Follow-up Designs

For purposes of this report, a working definition of "follow-up research studied" is offered. In general, the overall purposes of follow-up research are to determine if certain initially detected educational effects of schooling: (a) continue to play a role in improving an individual, or institution, (b) whether these effects diminish over the passage of time, or (c) are enhanced with the cumulative and eventual above "threshold" levels of educational exposure. Follow-up research also provides an opportunity to detect educational effects which take a longer period of time to emerge. Short-term perspectives (of perhaps even one school year) therefore are inadequate, since at that point, educational effects may still be latent. By underestimating educational effects, because of allowing too short a measurement time period for instructional features and events to exert an influence, misleading conclusions about the consequences or "pay-offs" of bilingual education (and similar large-scale, diverse innovations) may occur.

Follow-up research designs utilize data (baseline, or what can be usefully called "secondary data") from a prior study and subsequently collect additional data, which one may term "primary data." Such "follow-up" data are labeled primary because the investigator has maximum operational control over them, e.g., the secondary or baseline data is likely to be separated in time from the follow-up study. Follow-up studies can be planned at the same time (or practically at the same time) as the prior studies they seek to build on.

Follow-up studies may also consist of unplanned and essentially post hoc studies. Follow-up studies may collect primary data over an extended period of time, either on one or several occasions, as in longitudinal studies. Because of the nature of follow-up designs, it appears essential that a third type of data be collected, which can be termed "tertiary data." Such tertiary data may relate to baseline or secondary data, but probably more frequently it is collected because it appears related to primary or follow-up data. Tertiary data could include all types of data likely, in the investigator's judgment, to be useful in elucidating findings or in explaining unanticipated consequences.

The definition suggested above singles out several important characteristics of the follow-up research design. Exhibit 1, Typology of Follow-up and Related Designs, attempts to illustrate other attributes of follow-up designs by comparing and contrasting them with related designs. The typology in Exhibit 1 is based on theoretical and practical considerations and is not intended as a rigorous classification. However, to our knowledge, it does not exist elsewhere in the evaluation literature.

In Exhibit 1, two basic types of follow-up designs can be distinguished: Planned Follow-Up (No. 6) and Post Hoc Follow-Up (No. 7). Within these two basic types, follow-up study can be of other studies (No. 6a, or 7a) or of students, teachers, programs, etc. (No. 6b, or 7b). What obviously distinguishes the two basic types of follow-up designs is that the first is conceived and planned (and other provisions made, e.g., for tracking students) at the same time that an original study of program is planned. Under some conditions planned follow-up studies can take on the design attributes of an original study e.g., in terms of sample characteristics), and in some cases may even enhance the overall validity of both studies. In actual practice, one must acknowledge that the planning that takes place for a "planned follow-up" study is a matter of degree. Post Hoc follow-up studies are done as an afterthought, with little or no planning at the time of the conception of an original study.

Exhibit 1 is also meant to suggest that follow-up research designs generally lack most of the strengths of the experimental or quasi-experimental design, e.g., constraints imposed on the sensitivity of independent and dependent variables, adequacy of sampling, randomization and strength of inference. However, it is frequently the case that the more rigorous designs are extremely difficult to

EXHIBIT 1

TYPOLOGY OF FOLLOW-UP AND RELATED DESIGNS

Design Attributes

DESIGN TYPES	Adequacy of Sampling	Randomization	Overall Design Planning	Predominant Role of Theory	Prior Knowledge of Phenomenon	Long Term Time Perspective	Constraints on Baseline Variables (independent/secondary)	Constraints on Follow-up Variables (dependent/primary)	Prominence of Mediating Variables	Problem of Attrition	Determine Data Worthiness	Role of Data Analysis: Hypothesis Testing	Use of Tertiary Data: Elucidate Findings	Explain Unanticipated Consequences	Use of Retrospective Information	Strength of Inference
1. Experimental Research	x	x	x	x			x	x				x				x
2. Re-analysis of Data	x										x	x				
3. Continued Research	x										x	x				
4. Replication	x	x										x				x
5. Longitudinal	x	x	x	x		x	x	x	x	x		x				x
6. Planned Follow-up																
a. Of Other Studies			x	x	x	x		x	x	x	x	x	x	x		
b. Of Students, Teachers, Programs			x	x	x	x		x	x	x	x	x	x	x		
7. Post Hoc Follow-Up																
a. Of Other Studies				x	x	x		x	x	x	x	x	x	x	x	
b. Of Students, Teachers, Programs				x	x	x		x	x	x	x	x	x	x	x	
8. Qualitative Studies			x	x	x			x			x	x	x	x	x	

X- indicates an attribute associated with a design.

apply in school research, particularly research which focuses on language minority students in bilingual education settings.

In summary, the significant characteristics of follow-up research designs, and the conceptual, design and analytic issues which relate to these designs and their applications in bilingual education, may be stated as follows:

- Adequate and timely planning greatly enhances the validity and usefulness of follow-up studies, e.g., the critical problems of attrition can best be solved by planning ways of tracking or identifying subjects for follow-up measurement.
- Follow-up studies are derivative and in some respects retrospective, i.e., they are tied to prior studies or programs, and thus the investigator tends to lack adequate control. To compensate for the resulting design weaknesses, it is essential to use theory explicitly, to have a broad, multi-disciplinary knowledge of the phenomenon under investigation and to collect an excess of tertiary data which may be useful in analysis and interpretation.
- Follow-up designs impose maximum constraints only on the quality of follow-up or primary variables (dependent variables); therefore, the selection of follow-up variables and measures and the concurrent use of explicit theory for their basis becomes critical and requires more than usual ingenuity or creativity.
- Perhaps most significant of all characteristics is the extended time perspective (i.e., on treatment effects) of follow-up study. As a result of this extended time perspective, mediating variables are likely to have a prominent function, and more unanticipated consequences are likely to become apparent.
- The central problem of follow-up study is attrition. This problem can best be solved by timely planning. Effective but elaborate methodologies have been developed for identifying follow-up subjects but they tend to be somewhat costly. As Cochran and Cox said (1957): "The only complete solution to the problem of missing data is to not have any."
- Like most empirical research, hypothesis-testing plays a prominent role in follow-up studies.
- Follow-up study designs are well suited to exploring both the anticipated and unanticipated consequences, e.g., by collecting adequate tertiary data -- for instance by open-ended, ethnographic interviews of subjects.
- Follow-up designs are particularly well suited to examining learning thresholds which have been postulated to exist for a wide range of behaviors, including language acquisition, school learning, and training.

- Weaknesses inherent in follow-up designs must be partially compensated for by effective use of data analysis, e.g., an important role of data analysis lies in determining the worthiness of the data.

- Analysis of follow-up data confronts many of the same issues as in other types of research designs; however, the special features of follow-up data analysis are: (1) partial data treatments and (2) limitations in drawing inferences. Causal inference should be cautiously drawn from follow-up studies; at best, their findings probably should be considered correlational.

- Ethical considerations relating to the rights of subjects should figure prominently in planning and carrying out follow-up studies.

Selection of Follow-Up Variables: Measurement Issues

The selection of meaningful follow-up variables and sound sensitive measures is perhaps one of the most critical, as well as creative, tasks in planning a follow-up study. Generally, the investigator will have little or no control over baseline measurements and must use what is given. It was advocated that tertiary data be collected which bears on baseline measurements and which might possibly elucidate findings or unexpected consequences. Thus information on student characteristics may provide for reclassification of a sample population and thereby more effective analysis.

In essence, the crucial measurement issues in follow-up studies relate to: (a) the selection of relevant, sound and sensitive follow-up variables and measures, (b) the decisions about how to fill out the "knowledge context" of these variables by assembling easily gathered tertiary data, (c) and most importantly, posing the appropriate research question, one which is based on clearly elaborated theory and a body of accurate empirical knowledge.

Student Variables

In examining the long-term consequences of bilingual education for students one is basically interested in two things: the fluctuation of treatment effects over the long term (attenuation or its direct oppposite "snowballing", i.e., increasing effects) and other consequences which may be due to bilingual schooling. Treatment in bilingual education program consists of adding another language and culture to English monolingual education; thus, the distinctive variables involved in bilingual education are: L1 proficiency in the four language skills; L2 (English) proficiency; C1 enculturation (or socialization); and C2 enculturation.

Major consequences of bilingual education of interest to those doing follow-up study consist of academic achievement and other types of school attainment. Others, which relate to human development in general, consist of effects on students' personality (and related to this, attitude, motivation, self-esteem and value orientation), and career and life choices.

Teacher Variables

In examining the professional development of bilingual teachers from a long term perspective, follow-up studies should focus on the:

- Acquisition and maintenance of bilingual language proficiency;
- The proficiency levels required to teach in a particular language at a given grade level;
- Teacher instructional competence;
- Teacher knowledge of bilingual education theory and practices (and educational theory and practice in general);
- Attitudes toward L1 and L2, bilingual education, education, the teaching profession, and the local community; and finally,
- Teacher commitment to bilingual education as a profession.

Program Variables

A program focus would apply various implementation measures to the latter phase of the implementation process and to the onset of institutionalization within the local school system. The key implementation variable would consist of instruction or the instructional model being used. Secondary variables would consist of staff development, parent and community involvement, curriculum and material resources and program management. An additional, relevant variable would consist the capacity-building activities being carried out by the school system and program, to facilitate the process of institutionalization.

Large Scale Post Hoc Follow-up Studies

The feasibility and worthiness of conducting large scale post hoc follow-up studies of bilingual students, teachers or programs were examined in this paper. In bilingual education programs for language minority students, relevant baseline measurements may be quite diverse, limited in quantity or quality, or be difficult to obtain. The cooperation and active assistance of local school districts and programs seems essential in our view for conducting follow-up studies.

Much of the effort in conducting such follow-up studies would involve careful and lengthy identification of subjects with adequate baseline data, both for treatment and comparison groups, and subsequently tracking such individuals. Language minority students in bilingual programs can probably be expected to be even harder to track and locate, because of their greater mobility.

The authors conclude that post hoc studies can stand on their own merits as a legitimate and, perhaps, even preferable study design. A well-developed literature exists on the use of available data for social research purposes and such data tend to form a major part of post hoc studies.

Some writers have pointed out that many occasions exist when studies are not planned at the beginning of a program, or even where study procedures are not "in place" until a program is in progress or has actually terminated. The general problems of post hoc studies are well known and are particularly salient in the present context. Others find that the occurrence of these problems does not preclude the still-valuable use of post hoc follow-up designs.

Nonetheless, post hoc studies still need to deal with the limited quality of available data, the type of research questions being addressed, and their intrinsic value for examining meaningful fluctuations in program effects.

In general, current methodologies for locating follow-up subjects, which often employ creative, detective-like investigation techniques, are effective but often are somewhat costly.

The authors suggest the feasibility of conducting a long term post hoc follow-up study of students enrolled in bilingual education programs beginning in the 1966-67 school year (which were funded under Title I and Title III). Such students should have completed at least six years of elementary bilingual schooling. They would have graduated high school either this year or last and would be 18-20 years of age. They would also be in the process of making career and life choices. A second group (which might be even more amenable to long term follow-up study) would consist of students who entered Title VII bilingual programs in 1969-70, and who would now be eleventh graders.

Baseline measurements for long term student follow-up should include, if at all possible, measurements, or ratings of proficiency in L1 and L2 near the beginning and end of their bilingual schooling and measures of academic performance. A variety of additional "tertiary" data could be collected on student and program characterisitics.

Follow-up measurements would consist of language proficiency measures in L1 and L2 and a variety of school measures of success. Questionnaires and interviews could be used to obtain comprehensive information on students' social and cultural background, language skills and use, degree of enculturation (socialization) into minority group and mainstream culture and personality factors (attitude, motivation, self-esteem and value orientation). Information would also be elicited from students on their present and future life and career choices. A significant part of follow-up data gathering would consist of collecting additional information from students by means of ethnographic interviews, in order to probe some of the more subtle effects of their participation in bilingual programs. As for baseline measures, a variety of tertiary data would also be collected, e.g., from school documents, interviews with teachers and other school personnel, etc.

To our knowledge, no effort has been made to examine the long term professional development of bilingual teachers over an extended time perspective. Yet it is suggested that currently no other area in bilingual education would benefit more from a long term perspective. Both bilingual teacher education and in-service training would greatly benefit from long-term follow-up study.

A follow-up study of program implementation and institutionalization was suggested for one of the Part C Studies, "The Evaluation of the Classroom Instruction Component of the ESEA Title VII Program," which is basically a study of project implementation. Within one to two years after the completion of this study, there will be a need to focus a follow-up study on the later phase of implementing a bilingual program, and particularly on the processes of capacity building and institutionalization. This follow-up study could focus on the projects in the in-depth sample, from which comprehensive program characteristic and implementation data are being collected. Essentially this follow-up study would take the present Part C study one step further and answer critical questions about the process of institutionalization, i.e., how bilingual programs finally become incorporated in local school systems, and the trends in doing so.

Focused Follow-Up Studies of Bilingual Education Consequences for Students

A series of possible follow-up studies would use as baselines the data of the Significant Bilingual Instructional Features Study and its associated special studies. Follow-up data collection on the student populations involved could begin during the final phases of these studies or after they were completed. However, planning such studies should begin as soon as feasible. Each follow-up study would have different needs for follow-up measures, observational protocols and other data gathering procedures.

The authors believe strongly that the soon-to-begin bilingual instructional features studies should be supported by small, carefully designed follow-up substudies of substantive issues and problems in bilingual education. Some of these follow-up studies could, perhaps, be independent of the antecedent studies, and would resemble replications. However, they would all have in common an extension of "time perspective." They would both contribute to knowledge, and increase the credibility and acceptability of the instructional features studies as a whole.

These focused follow-up studies could include:

- Continued Research on Instructional Features
- Language Acquisition Follow-up
- Language Use Follow-up
- Cognitive Development Follow-up
- Case Study Follow-up of Personality Factors
- Case Study Follow-up of Students' Performance in All-English Speaking Classrooms

Each of these proposed focused follow-up studies is discussed in the report in terms of follow-up, baseline and tertiary data requirements; research questions; design and methodological issues and problems; and the meaning of expected categories of results.

Conclusion

There is a definite place for follow-up studies in investigating the long-term, enduring and even unanticipated effects of bilingual schooling. However, these investigations will need to be based on a recognition of their value to educational audiences, relative merits of research issues, attention to the problem of attrition, proper selection of dependent measures, and the merits of planned follow-up studies. It is acknowledged that post hoc follow-up research also has merit, if that form is the more realistic to conduct. Follow-up studies also will benefit a great deal by: (a) using comprehensive and sound baseline measures, from which trends in educational effects can be appraised, and (b) annexing tertiary data as available to the follow-up data base so that trends in observed effects can be interpreted, explored in rich detail, and finally plausibly explained.

Bilingual education, as a highly diverse (perhaps the most diverse educational innovation that ever has been mounted as a federal funding initiative), could benefit a great deal from follow-up studies. If for nothing else, shorter term effects attributable to bilingual program participation may be more difficult to detect, given the wide diversity of students, programs, available assessment instruments and local contexts. Thus, the preliminary planning of future, long term research should be considered, so that the benefits and educational consequences of significant features present in bilingual settings will not be underestimated and overlooked.

1.0 INTRODUCTION

1.1 Purpose of Planning Paper

The proposal for the Bilingual Instructional Features Planning Study called for a final task, Task 7, which would "indicate options for periodic follow-up." It also stated that "the purposes to be served by a follow-up study" would be indicated and that "a number of alternatives for tracking samples of children be considered." This Task also was described in the RFP (79-0026) as "indicate advantages/disadvantages of alternative low-cost strategies for continuing a modified version of the data collection . . . on a periodic basis."

The proposal and RFP suggested that such follow-up options be described for the study or studies deriving from the Bilingual Instructional Features Planning Study. There is no question that the major descriptive study and the three special studies do offer the primary opportunity for follow-up; consequently, this planning report examines the issues and options for follow-up study of these four studies.

Knowledge of the revised DHEW Bilingual Education Research Plan (1979), the RFPs for the planning study and subsequent studies, and previous reports in this series of planning papers is assumed (See Foreword for descriptions of the contents of planning papers).

Since major results from the large scale descriptive study of instructional features and the three special studies are needed early to satisfy Congressional mandates, these studies will have to report on fairly short-term effects (after one year with subsequent reporting during a third year). The importance of these studies indicates that they should have long term or follow-up component if their results are to achieve full credibility. In preparing this paper, its authors became convinced there are other foci and areas for follow-up which need to be examined in a planning paper dealing with follow-up studies of bilingual education.

We believe it worthwhile to examine the broader areas in which follow-up studies may be carried out. We suggest that these areas are:

(a) Selected Part C studies

(b) Follow-up of other individual studies which offer unusual promise or opportunity

(c) Follow-up of students, teachers and programs, not merely follow-up of existing studies

We also think it timely to include a more comprehensive review of follow-up methodology in this planning paper, particularly as it may be applied in bilingual education. In adopting this broader view of follow-up research, we believe we are more properly addressing the concerns of not only the National Institute of Education and the Part C Coordinating Committee, but in particular organizations such as the National Center for Bilingual Research. One of that organization's major activities will emphasize development of longitudinal studies, data generation activities and its role as a central focus for bilingual education in the nation.

This paper is organized into nine sections. Sections 1, 2 and 3 introduce the major issues of follow-up study in bilingual education and tentatively indicate broad options. The need for follow-up research in bilingual education is assessed, and an attempt is made to identify the studies or areas which are potentially suitable for follow-up research. Significant options for follow-up research, in terms of general foci or units of analysis are also discussed. Finally, an overview of the nature of follow-up research is given.

Sections 4 and 5 deal with the conceptual, methodological and analytic issues in planning and constructing follow-up designs for bilingual education. The authors found it essential to treat these issues at length here because they apparently have never been brought together and treated as a whole within the context of bilingual education.

Section 6 deals with the measurement and data-gathering issues in follow-up research in bilingual education, and emphasizes the selection of variables for use in follow-up studies.

Section 7 examines the feasibility of conducting large scale yet post hoc (relatively unplanned or unanticipated) follow-up studies of: (a) long term consequences of bilingual education for students, (b) long term professional development of bilingual teachers, and (c) the final phases of bilingual program implementation or institutionalization. Though certain weaknesses are noted in such post hoc follow-up designs, the authors of this paper recommend that some of these studies be considered further. These recommendations, therefore, are elaborated as preliminary versions of follow-up study designs.

Section 8 examines the feasibility of conducting focused studies of the consequences of bilingual educaton for students, based on the future Significant Bilingual Instructional Features Study and its three related special studies. Preliminary, focused follow-up study designs are described for continued research on instructional features and their consequences, language acquisition, language use, cognitive development, attitudinal and other personality factors, and school-related skills and behaviors in all-English speaking classrooms.

Section 9 concludes the Planning Paper with a summary appraisal of the merits of conducting follow-up research in bilingual education.

1.2 Need for Conducting Follow-up Studies

A long term perspective or interest in the enduring effects of education in the United States could almost be called a national or cultural blind spot. Rare indeed are the studies that follow children through the school system and into college or adult life. (Two such do come to mind: Project Talent, and the National Longitudinal Study of the High School Class of 1972.) In the case of the diverse innovation of bilingual education, a blind spot for long term perspective on education and human development again appears to exist. Research and policy evaluation continue to insist on the quick, yet perhaps unfair and misleading demonstration of achievement over a relatively short period of time. The conditions for long-term research are present, especially when one considers that federally funded bilingual education has been around since 1967 through ESEA Title I and III, and with much greater intensity since 1968 with the passage of the Bilingual Education Act. As an example of a specific research condition, one can note that in Spring 1980 a group of some 40 high school graduates in Las Cruces, New Mexico, represented an intact group throughout almost 12 years of bilingual schooling. Surely even in a post hoc fashion such a group would be worthy of study.

One would hope that such a radical innovation in American education would have been under examination in order to determine the long term consequences for students, teachers and programs. This has not been the case. Thus, the long term consequences of bilingual education remain largely unexamined.

Tucker (1979) finds that the lack of longitudinal studies is one of the most severe problems in obtaining hard empirical data about the benefits and consequences of bilingual education:

> Critical, empirical, and longitudinal assessments have been particularly lacking. The paucity of such studies can be attributed to a variety of sources such as the transience of potential researchers, the unwillingness or inability of funding agencies to commit funds for a multiyear or even decade long project, . . . and the unfortunate tendency for administrators to regard initial results as the major criterion for continuing or terminating a proposed lengthy project.

St. Pierre's recent (1980) highly useful article on longitudinal field studies expressed similar concerns for there being "powerful arguments for using longitudinal studies whenever possible."

Such longitudinal studies of bilingual education are probably infeasible at the moment, although having the results of well-conceived follow-up studies of existing studies and programs would be a useful interim step. Follow-up studies employing a long term perspective are needed to confirm and further explore largely academic and suggestive accounts of the effects of bilingualism and bilingual education on: language acquisition, cognition, problem solving, educational achievement, personality and character formation, attitudes, career and life choices of individuals (Leopold, 1939-49; Peal, 1962; Diebold, 1968; Gezi, 1975). There is a specific need to confirm the existence and consequences of language learning thresholds for dominant and second languages (i.e., L1 and L2). Information is also needed on exit criteria and the subsequent placement of bilingual education students in school system mainstream classes. Furthermore, information is needed on the type of support students require (if any) after they leave bilingual education classrooms.

Above all, a need exists to explore further all the meaningful consequences -- anticipated and unanticipated; positive and negative -- of students' participation in types of bilingual education programs.

On another plane, what has been said of the long term development of the student can also appropriately be said for the bilingual teacher. There is a need for a long term perspective on the training and professional development of the teacher. Systematic knowledge on the long term development of bilingual teachers is a significant gap in bilingual education and in-service training.

Finally, the long term perspective, which could be provided by various forms of follow-up study, needs to be applied to programs. Program implementation and particularly program institutionalization studies are needed. That is, a long term follow-up would be able to provide a better understanding of how and to what extent programs have been incorporated into local school systems, once external (usually federal) funding has terminated.

1.3 Overview of Follow-Up Designs

For purposes of this report, a working definition of "follow-up research studies," can be offered. In general, the overall purposes of follow-up research are to determine if certain initially detected educational effects of schooling: (a) continue to play a role in improving an individual, or institution, (b) whether these effects diminish over the passage of time, or (c) are enhanced with the cumulative and eventual above "threshold" levels of educational exposure. Follow-up research also provides an opportunity to detect educational effects which take a longer period of time to emerge. Short-term perspectives (of perhaps even one school year), therefore, are inadequate since at that point, educational effects may be latent. By underestimating educational effects, because of allowing too short a measurement time period for intructional features and events to exert an influence, misleading conclusions about the consequences or "pay-off" of bilingual education (and similar large-scale, diverse innovations) may occur.

Follow-up research designs utilize data (baseline, or what can be usefully called "secondary data") from a prior study and subsequently collect additional data, which one may term "primary data." Such "follow-up" data are labeled primary because the investigator has maximum operational control over them, e.g., the secondary or baseline data is likely to be separate in time from the follow-up study.

Follow-up studies can be planned at the same time (or practically at the same time) as the prior studies they seek to build on. Follow-up studies may also consist of unplanned and essentially post hoc studies. Follow-up studies may collect primary data over an extended period of time, either on one or several occasions, as in longitudinal studies. Because of the nature of follow-up designs, it appears essential that a third type of data be collected, which can be termed "tertiary data." Such tertiary data may relate to baseline or secondary data, but probably more frequently it is collected because it appears related to primary or follow-up data. Tertiary data could include all types of data likely, in the investigator's judgment, to be useful in elucidating findings or in explaining unanticipated consequences (Messick, 1970).

The definition suggested above singles out several important characteristics of the follow-up research design. Exhibit 1, Typology of Follow-up and

Related Designs, attempts to illustrate other attributes of follow-up designs by comparing and contrasting them with related designs. The typology in Exhibit 1 is based on theoretical and practical considerations and is not intended as a rigorous classification. However, to our knowledge, it does not exist elsewhere in the evaluation literature.

In Exhibit 1, two basic types of follow-up designs can be distinguished: Planned Follow-Up (No. 6) and Post Hoc Follow-Up (No. 7). Within these two basic types, follow-up study can be of other studies (No. 6a, or 7a) or of students, teachers, programs, etc. (No. 6b, or 7b). What obviously distinguishes the two basic types of follow-up designs is that the first is conceived and planned (and other provisions made, e.g., for tracking students) at the same time that an original study or program is planned. Under some conditions, planned follow-up studies can take on the design attributes of an original study (e.g., in terms of samples) and in some cases may even enhance the overall validity of both studies. In actual practice, one must acknowledge that the planning that takes place for a "planned follow-up" study is a matter of degree. Post Hoc follow-up studies are done as an afterthought, with little or no planning at the time of the conception of an original study.

Exhibit 1 is also meant to suggest that follow-up research designs generally lack most of the strengths of the experimental or quasi-experimental design, e.g., constraints imposed on the sensitivity of independent and dependent variables, adequacy of sampling, randomization and strength of inference. However, it is frequently the case that the more rigorous designs are extremely difficult to apply in school research, particularly research which focuses on language minority students in bilingual education settings.

The significant characteristics of follow-up research designs, and the conceptual, design and analytic issues which relate to these designs and their applications in bilingual education, are summarized below. Methodological issues affecting the use of follow-up designs are treated at length in section 4.

EXHIBIT 1

TYPOLOGY OF FOLLOW-UP AND RELATED DESIGNS

Design Attributes

DESIGN TYPES	Adequacy of Sampling	Randomization	Overall Design Planning	Predominant Role of Theory	Prior Knowledge of Phenomenon	Long Term Time Perspective	Constraints on Baseline Variables (independent/secondary)	Constraints on Follow-up Variables (dependent/primary)	Prominence of Mediating Variables	Problem of Attrition	Role of Data Analysis: Determine Data Worthiness	Hypothesis Testing	Use of Tertiary Data: Elucidate Findings	Explain Unanticipated Consequences	Use of Retrospective Information	Strength of Inference
1. Experimental Research	x	x	x	x			x	x				x				x
2. Re-analysis of Data	x										x	x				
3. Continued Research	x										x	x				
4. Replication	x	x										x				x
5. Longitudinal	x	x	x	x		x	x	x	x	x		x				x
6. Planned Follow-up a. Of Other Studies			x	x	x	x		x	x	x	x	x	x	x		
b. Of Students, Teachers, Programs			x	x	x	x		x	x	x	x	x	x	x		
7. Post Hoc Follow-Up a. Of Other Studies				x	x	x		x	x	x	x	x	x	x	x	
b. Of Students, Teachers, Programs				x	x	x		x	x	x	x	x	x	x	x	
8. Qualitative Studies			x	x	x			x			x	x	x	x	x	

[1] X - indicates an attribute associated with a design.

- Adequate and timely planning greatly enhances the validity and usefulness of follow-up studies, e.g., the critical problem of attrition can best be solved by planning ways of tracking or identifying subjects for follow-up measurement.

- Follow-up studies are derivative and in some respects retrospective, i.e., they are tied to prior studies or programs, and thus the investigator tends to lack adequate control. To compensate for the resulting design weaknesses, it is essential to use theory explicitly, to have a broad, multi-disciplinary knowledge of the phenomenon under investigation and to collect an excess of tertiary data which may be useful in analysis and interpretation.

- Follow-up designs impose maximum constraints only on the quality of follow-up or primary variables (dependent variables); therefore, the selection of follow-up variables and measures and the concurrent use of explicit theory for their basis becomes critical and requires more than usual ingenuity or creativity.

- Perhaps most significant of all characteristics is the extended time perspective (i.e., on treatment effects) of follow-up study. As a result of this extended time perspective, mediating variables are likely to have a prominent function, and more unanticipated consequences are likely to become apparent.

- The central problem of follow-up study is attrition. This problem can best be solved by timely planning. Effective but elaborate methodologies have been developed for identifying follow-up subjects but they tend to be somewhat costly. As Cochran and Cox said (1957, cited in Morell 1979): "The only complete solution to the problem of missing data is to not have any."

- Like most empirical research, hypothesis-testing plays a prominent role in follow-up studies.

- Follow-up study designs are well suited to exploring both the anticipated and unanticipated consequences of treatment. Such designs must, therefore, prepare for such unanticipated consequences, e.g., by collecting adequate tertiary data -- for instance by open-ended, ethnographic* interviews of subjects.

- Follow-up designs are particularly well suited to examining learning thresholds which have been postulated to exist for a wide range of behaviors, including language acquisition, school learning, and training.

- Weaknesses inherent in follow-up designs must be partially compensated for by effective use of data analysis, e.g., an important role of data analysis lies in determining the worthiness of the data.

---

*The reader is referred to Planning Paper 6 in this series (Rudes, Goldsamt and Cervenka, 1980) for a thorough exposition of issues in applying ethnographic methodology to bilingual classroom settings.

- Analysis of follow-up data confronts many of the same issues as in other types of research designs; however, the special features of follow-up data analysis are: (1) partial data treatments and (2) limitations in drawing inferences. Causal inference should be cautiously drawn from follow-up studies; at best, their findings probably should be considered correlational.

- Ethical considerations relating to the rights of subjects should figure prominently in planning and carrying out follow-up studies.

2.0 OPTIONS OR AREAS OF FOCUS FOR CONDUCTING FOLLOW-UP INVESTIGATIONS

The proposal for the Bilingual Instructional Features Planning Study (Phase I) suggested that follow-up options be explored for the study or studies deriving from the Significant Instructional Features in Bilingual Education Study (Phase II). However, in preparing this paper, its authors became convinced that still other studies and activities -- including some which focused on units of analysis other than the student -- could profitably be examined in a planning paper dealing with follow-up studies in bilingual education.

This report, therefore, treats the major descriptive study of Phase II and its accompanying special studies as the primary current opportunity for follow-up study in bilingual education. However, the authors found it worthwhile to examine the broader area in which follow-up studies in bilingual education may be carried out, viz:

a. Selected Part C Studies.

b. Other individual studies which offer unusual promise or opportunity.

c. Current new research on students, teachers and programs (i.e., not merely follow-ups of prior studies).

The Part C Studies addressed the national need for bilingual education, the improvement in effectiveness of services for students and the improvement in program management and operation. Several of these studies, therefore, provide a basis for follow-up, or at least access to data which might prove useful in follow-up studies.

2.1 Significant Instructional Features Study (B-1)

This major three-year descriptive study seeks to collect information on instructional features in bilingual education programs and on their consequences for students (as well as contextual data) during its first phase. It will attempt to confirm linkage between features and consequences, primarily during its second phase. A third phase will be geared to providing recommendations for improving instructional practices in bilingual education on the basis of the study findings. During the first year, in particular, a large quantity of qualitative data will be gathered on students, teachers, schools and communities.

2.2 Special Studies Associated with B-1

Language Learning in Bilingual Instruction (Study 1) will seek to determine what instructional practices foster school-related skills in the second language, and also how student language characteristics interact with bilingual instructional practices to affect acquisition of language skills. The data to be collected include: the four language skills (listening comprehension, speaking, reading, writing) with emphasis on literacy data; language use data; cognitive aspects of language development (thinking, remembering, problem-solving, etc.); and parental questionnaire data. Sites having variations in instructional practices will be used. The project is intended to last two to four years.

Language Attitudes of Parents, Students and Teachers in Bilingual Classrooms (Study 2) will examine how parents, teachers and students' attitudes toward language, culture and bilingual instruction and language influences behaviors and instructional features in bilingual classrooms. This study also deals with how the attitudes of parents, teachers, and students affect each other and potentially might change over time. The principal data to be collected are language attitudes and classroom behavioral data. Data collection sites may consist of school and/or schools. The study will examine variations in programs employed by language minority schools.

Instructional Practices in Non-Public Schools (Study 3) will be a case study of non-public schools with bilingual education programs. The unit of analysis will be the school and its program features. Observational data will be collected on classroom and school practices. Descriptions of schools by school officials will also be used. The study will examine the relationship of the school setting to the community, teacher characteristics, and those of students and their families.

2.3 Part C Studies

The Part C Studies which address the national need for bilingual education (see the revised HEW Research Plan for Bilingual Education, 1979), provide a basis for conducting follow-up studies or at least data which might be used in follow-up studies. These would include:

- Survey of Children With Limited English Speaking Proficiency and the Services Which They Receive (termed A-1 in the Research Plan);
- Update of Current National Data on Teacher Language Skills (A-6); and
- Study of the Need for Bilingual Education in Puerto Rico (A-7).

The Part C Studies which address the improvement in the effectiveness of services for students would appear to offer abundant opportunities for determining the long term trends and effects of instruction. Aside from B-1, these would include:

- A Study of Title VII Inservice Training Needs Assessment Processes and Activities (B-2);
- A Study of Parental Involvement in Bilingual Education Programs (B-3);
- Development and Dissemination of Instructional Models (B-4);
- Development and Dissemination of Evaluation and Data Gathering Models (B-5);
- Development of Evaluation Models for ESEA Title VII Bilingual Education Projects (B-5.1);
- Refinement and Dissemination of Standards for Identifying Students to Receive Bilingual Education (B-6); and
- Support of Field-initiated Proposal for Research on Bilingual Education (B-7).

The Part C Studies which address the area of improvement in program management and operations also would appear to offer some opportunity for determing if certain sustaining effects have occurred and/or certain benefits of bilingual settings have manifested themselves over the long-run, rather than not appearing as a short-term effect. Follow-up research, therefore, could be quite useful. They would include:

- Evaluation of the Classroom Instruction Component of the Title VII Bilingual Education Program (C-1);
- Development and Installation of a Title VII Program Data System (C-2);
- Impact of the Title VII-Funded Training Programs Operating by Institutions of Higher Education (C-5); and
- Study of the Impact of Title VII Capacity Building Trust (C-6).

2.4 Promising Follow-Up Opportunities

Given the fact that publically funded bilingual education has existed since 1967, the authors contend that there are substantial opportunities for follow-up research on: (a) prior or on-going studies and (b) on current research dealing with teachers and programs. They argue that even in a *post hoc* fashion such studies can be worthwhile.

These studies could be follow-ups of students who have participated in bilingual education programs over a long period of time, e.g., the programs in Dade County, Florida; Las Cruces, New Mexico; or Laredo, Texas. These would be *post hoc* studies using available baseline data such as test scores, evaluation data, report cards and other measures of school success. Research questions and follow-up measures would be formulated based on a knowledge of the setting and issues that are valid to address using that setting (See Sections 3 and 6) of this Planning Paper.

Other follow-up research might consist of following up selected studies of students, teachers or programs, e.g., the Redwood City, California Bilingual Program (Cohen *et al*., 1975; Cohen and Laosa, 1976); the Douglas, Arizona, Bilingual Program (Soldate and Mishra, 1978); and perhaps the A.I.R. Study (1977, 1978).* Other studies funded by NIE which are perhaps worthy of follow-up might include: Community and School Interactions in Bilingual Education, Assessing the Language Proficiency of Bilingual Persons, and In-Service Teacher Training for Bilingual Education.

---

*In response to criticism, A.I.R. conducted a follow-up study which encountered additional problems. A.I.R. identified a follow-up group to be tested in Fall 1976, a full year after the pretest. Because of funding problems only a subset of classrooms could be considered. These contained 277 Title VII and 71 comparison children in Grade 2. But there were only 44 children with complete data in sites with comparison children, and only 15 of these children could have been LEP. Thus, the sample sizes used in follow-on analyses were too small to provide meaningful conclusions.

3.0 FOLLOW-UP STUDY RESEARCH FOCI

The authors argue that there are three primary foci or aspects of current bilingual education which could be suitably examined by means of follow-up research designs: (a) student focus -- long term consequences; (b) teacher focus -- long term professional development; and (c) program focus -- implementation and institutionalization. Each of these foci represent the principal levels of analysis needed for future follow-up studies. Other foci are possible, such as combinations of these foci, e.g., to what extent teacher gains lead to student gains over time. However, these appear to be justified by current circumstances in bilingual education. Each of these areas will be discussed in turn below.

3.1 Student Focus: Long Term Consequences

The primary focus for long term follow-up studies in bilingual education clearly is the student. The nature of his/her experience in bilingual education classrooms, what is learned in such classrooms, and are what the consequences which result from participation in bilingual education programs are major issues. In fact, the importance of the student focus follows from many policy, programmatic and local practical concerns. The focus on the student is important because progress or positive consequences are expected at this level in education; affecting change is a prescribed role of education in U.S. society.

This focus is also justified in a follow-up study because of the greater likelihood that useable, valid measurements of student ability can be carried out (i.e., as compared with other possible foci). In some situations it may be feasible to carry out follow-up studies on intact classes of students (or grades, as in various cross-sectional cohort or longitudinal study designs). A set of illustrative research questions which could guide inquiry on long term student consequences is offered in Exhibit 2.

EXHIBIT 2

POTENTIAL RESEARCH AREAS TO INVESTIGATE DEALING WITH THE LONG TERM CONSEQUENCES OF BILINGUAL EDUCATION ON STUDENTS

1. What school-related skills and behaviors are needed by students in bilingual education programs to function effectively in all English speaking classrooms?

2. What factors are the best predictors of academic success in mainstream classrooms for students in bilingual education programs?

3. What factors should be considered in the placement of bilingual students into mainstream classrooms?

4. What types of follow-up treatment, if any, are needed by bilingual students who are placed in mainstream classrooms?

5. To what extent do English language skills predict effective functioning in all English speaking classrooms?

6. What effect does bilingual education have on student attitudes, motivation, self-esteem and value orientation?

7. What effect does bilingualism have on cognition, particularly problem solving?

8. What consequence does bilingual education have on student life and career choices?

9. What factors account for the persistence of bilingualism over time?

10. What factors account for the persistence of bicultural socialization over time?

11. What is the relationship between bilingualism and loyalty to L1 and L2, community, state and nation?

12. What are the long term consequences for L1 and L2 reading ability, of students learning to read first in their native language?

13. What are the long term consequences for L1 and L2 reading ability, of students learning to read first in English (L2)?

14. What are the long term consequences for L1 and L2 reading ability, of students learning to read L1 and L2 at the same time?

15. What oral-aural language threshold level does a student have to reach in L1, in order for his/her subsequent development in that language to proceed in a normal pattern?

16. What oral-aural language threshold level does a student have to reach in L1, in order for his/her subsequent development in L2 to proceed in a normal pattern?

3.2 Teacher Focus: Long Term Professional Development

In terms of policy, programmatic and local practical concerns, a study focus on bilingual teachers must be regarded as secondary to that of students. However, there is perhaps no other area in bilingual education where a long term perspective would be more useful. Findings on the long term professional development of teachers, including their graduate and undergraduate education, their service as teacher aides or paraprofessionals, their practicums, and their experience in bilingual classrooms, could provide a basis for reforming bilingual teacher education and inservice training. Exhibit 3 presents a set of illustrative research questions which could serve to guide inquiry on the long term development of teachers.

EXHIBIT 3

POTENTIAL RESEARCH AREAS DEALING WITH THE LONG TERM DEVELOPMENT OF TEACHERS

1. What level of proficiency in a second language is needed in order to teach effectively in that language at a given grade level?
2. How do teachers acquire proficiency in two or more languages?
3. What factors account for teachers' instructional competence in bilingual classrooms?
4. What factors account for the "knowledge background" that bilingual teachers need, to be effective teachers?
5. What are the relative contributions of: social-cultural background, formal education, practicum, service as an aide or paraprofessional, teaching experience, inservice training and other factors to the overall development of a bilingual teacher?
6. What key factors most differentiate teachers who started as bilingual teacher aides or paraprofessionals, and teachers who obtained teaching credentials without this experience?
7. What types of attitude changes occur in bilingual teachers over a long term period of time? How do these changes affect their roles as bilingual teachers?
8. What factors determine a teacher's professional commitment to education as a profession, and to bilingual education in particular?
9. What factors appear to influence a teacher's choice of bilingual education as a career?

3.3 Program Focus: Implementation and Institutionalization

A current major study, "The Evaluation of the Classroom Instructional Component of the ESEA Title VII Program," funded by the Education Department,* focuses on implementation and the program as the major unit of analysis. To some extent, the study deals with capacity-building. However, there still is a greater need to focus on the latter implementation stages of bilingual programs, and on how such programs become institutionalized in local school systems. Exhibit 4 therefore presents a series of illustrative research questions which could serve to guide inquiry in this area.

EXHIBIT 4

POTENTIAL RESEARCH DEALING WITH LONG TERM PROGRAM IMPLEMENTATION AND INSTITUTIONALIZATION

1. What factors influence the latter phases of program implementation?
2. What processes of implementation appear more successful than others?
3. What specific strategies or methods appear to be more successful in implementing bilingual programs during their final phases?
4. To what extent are programs and their components (instruction, staff development, parent and community involvement, and management) implemented before they are institutionalized?
5. What are the most effective ways to building capacity for continuing bilingual education programs?
6. What criteria should be utilized in order to determine whether a program is adequately institutionalized in a local school district, so that external funding needs no longer exist?
8. How can the extent of program implementation be most effectively measured?
9. How can the extent of program institutionalization be most effectively measured?

---

*Awarded to Development Associates, Inc. and its subcontractor, Abt Associates, Inc.

4.0 METHODOLOGICAL ISSUES AFFECTING THE USE OF FOLLOW-UP DESIGNS

This chapter will deal with a variety of issues which affect the design of, and interpretation of findings from, follow-up studies. For each of these issues we will attempt to indicate its ramifications and implications for bilingual education research, and present some of the plausible alternatives to dealing with the issue. This chapter should also be read as an extension of the section presented earlier (1.3) and its associated Exhibit 1, Typology of Follow-Up and Related Designs.

4.1 General Considerations

It should be kept firmly in mind that follow-up studies, because of their very nature, are derivative and retrospective in form. They arise out of a need to assess if certain earlier-detected program effects, for example, continue to affect students. Thus follow-up studies are inherently associated with the baseline data levels collected through a prior research investigation. Thus, any deficiencies which are part of the original design tend to come along and be part of the follow-up study. The latter therefore is likely to be using the same dependent measures of student progress, the same comparison groups, and the same sampling design with respect to the variables which were used to establish representative samples, intact bilingual classrooms, selection of bilingual projects, etc. Weaknesses in the validity or reliability of the dependent measures, pre-program participation differences in comparison groups, and inherent sampling biases due to large sampling errors therefore are difficult to compensate for in the "piggy-back" follow-up design.

However, to some extent, the follow-up design can compensate for certain earlier deficiencies. The interpretation of follow-up data based on dependent measures can be shored up by the use of an explicit theoretical rationale, one which clarifies the pattern of trends in data occurring over time, and relates it to the theory and its corollaries. Such a theory may deal with maturational changes, the process of learning called "enculturation," gains in L2 surpassing a specified threshold level of linguistic competence in L1, etc.

Whatever the theoretical position chosen to explain the pattern of data (and preferably a theory adopted prior to the examination of data so that hypothesis testing efforts can occur), using explicit theory is much better than an attempt which is ungrounded in theory and which tries to explain in a *post hoc* fashion the trends in the obtained follow-up data.

For much the same purpose, having a comprehensive, multidisciplinary knowledge of the study comparison groups and units of analysis and their local contexts -- both at the onset of the study and as the study progresses -- is invaluable. Having this knowledge can go a long way toward ruling out alternative, competing explanations for program effects, and for accounting for unanticipated consequences or trends in data which may occur. Having this comprehensive knowledge at the onset of the follow-up study will help to understand in what ways the comparison groups are similar, and in what ways they are systematically different, especially on factors or dimensions which are apt to confound and interact with the dependent measures. Such knowledge should not stop or become irrelevant as the follow-up study begins and its data are collected. Bilingual education is highly dynamic rather than static. At any time federal, state or local regulations are apt to affect programs (certification requirements, for example). Key program administrative and teaching personnel may leave the local setting, and different levels of staff may become the key actors. Local bilingual populations may become more visible and vocal and/or exist in greater numbers due to emigration into the attendance area.

Such factors have been recognized as important to be aware of by various investigators (Fox *et al*., 1976 or Glass and Maguire, 1977, who considered "intervention factors" accounting for a dynamic time series trend of data points). Dealing with these issues analytically is discussed in Chapter 5.0. For the time being, it is recommended that investigators conducting follow-up research would be well-advised to keep abreast of events in time which could potentially affect their data. A simple log or diary of such events is perhaps a minimal requirement to record the occurrence of these events, and their ramifications.

4.2 Unanticipated Consequences

While certain follow-up studies have the goal of determining if the initial study's findings of short-term (perhaps one school year) effects or gains will continue (or even become greater due to cumulative educational experiences), other studies with a follow-up component may find that only in the latter phase does a program impact occur, or that other forms of program-attributable effects occur.

Such unanticipated consequences are likely to occur as sufficient time passes, so that intervening events occur between the baseline and follow-up measurement. Such intervening events thus become relatively more likely to have played a role in causing the follow-up measurement than the bilingual program, the inservice training, or similar aspects of bilingual program operations.

Yet unanticipated consequences do not have to remain this way, if the data collection scope can be expanded so that subtle programmatic effects can be detected. Thus student, teacher or administrator ethnographic interview data can be gathered to explore the full context of the bilingual setting. Such ethnographic interviews represent a form of tertiary data and can be invaluable for generating hypotheses which relate baseline measurements in one area (such as student self-esteem) to follow-up measurements in other areas (such as reading achievement). Thus ethnographic interviews, like maintaining a log of events affecting bilingual program operations during the full baseline and follow-up study period, are forms of data gathering strategies which accommodate and can deal with some unanticipated consequences or findings.

Carrying this point further, if such ethnographic data were gathered from well-chosen comparison groups, (such as elementary and high school students in bilingual education) then the opportunity would exist for assessing if the unanticipated consequences are restricted to a particular educational level or are wider-based. This also would help to determine if the unanticipated consequences appear related to certain types of programs and not others.

4.3 Learning Thresholds

This issue refers to the possibility of interactions between educational programs and chid input factors not having been considered in designing evaluation studies. Such an omission was pointed out by Cummins (1979) who stated that "...one must consider the dynamics of the bilingual child's interaction with this educational environment if any answer is to be found to the central question of whether or not the academic progress of children of limited English speaking ability will be promoted more effectively if initial instruction is in their L1." With this central point in mind, Cummins put forth the thesis that children must attain threshold levels of linguistic competence if being bilingual is to become a benefit and influence cognitive growth. Attaining such a threshold level also will avoid cognitive deficits.

Thus maintaining L1 skills is likely to be a prerequisite for attaining higher threshold levels of linguistic competence, and correspondingly, cognitive benefits for minority group children. In this form linguistic competence level is likely to serve as an intervening variable in mediating effects of bilingualism on cognitive and academic development.

The research studies cited by Cummins (1979) in postulating both the linguistic competence threshold concept, and later in the same paper, an interaction model of bilingual education, were suggested to be applicable to explaining cognitive behaviors of various forms. The model is likely to be a useful explanatory tool for dealing with follow-up study data trends. In fact, in a similar vein, Cummins stated that "...educators take account of the diversity of input characteristics of their students...evaluations must follow a 'planned variations' approach in order to find the optimum blends of Input and Treatment characteristics under sociocultural conditions. Failure to take account of possible Input x Treatment interactions is likely to result in uninterpretable data."

Thus it may be that the interaction model postulated by Cummins, and the form of additive bilingualism (literacy in both L1 and L2) he considered optimal for promoting cognitive and academic potential, represents one way to resolve uninterpretable follow-up findings. Those with wide time spans between baseline and follow-up measurements may be aided by the tertiary data sets mentioned earlier and the use of an explicit modeling approach such as the threshold competence concept and the interaction model.

4.4 Use of Comparison Groups

Follow-up studies should use comparison groups as a basis for understanding if data trends in the bilingual student group (the "treatment") are truly associated with participation in the program, rather than being due to maturational changes or other factors. Using a group of students who are not in the program, yet who are comparable or near-equivalent in most other relevant characteristics (age, sex, grade level, language proficiency, etc.) will therefore help in ruling out spurious treatment effects.

However, this is easier said than done. Randomized group assignment and having treatment-control groups would equalize both groups at the onset of research, yet is difficult to achieve due to administrative regulations and "denial of service" issues. Also, it is difficult due to the very nature of federal funding policy, e.g., Title VII grantees are to serve those whose limited English speaking needs warrant being in the program. Very few children not in the program therefore are likely to be comparable to these in the program within a school system. An exception to this pattern might be those in urban school systems on a waiting list for admission, due to limited classroom sizes and teaching staff. However, the likelihood of having a "pure" and equivalent intact (i.e., nonrandomized) comparison group for a treatment group, yet not being in the bilingual program is a near-impossibility. Comparison groups are more likely to be composed of children who systematically differ in certain ways from the treatment group which is the focus of the follow-up study. At that point, study efforts should conduct information-gathering activities to fully understand in what antecedent ways the treatment and comparison group differ. Such design and statistical approaches will thus attempt to unconfound or partition the effects of these differences from observed trends in score differences between the two groups.

Not only may the demographic and linguistic characteristics of treatment and comparison groups differ, but the latter may be obtaining some form of educational services in the same school setting. Such instruction may not be funded by Title VII but by other federal sources such as Indian Education, Indochinese Refugee Assistance, Cuban Emergency Aid, etc. Since the comparison group also may gain in cognitive, personality, career aspiration levels over time, perhaps due to the instructional experiences they are having, it becomes necessary to fully understand the nature of what instruction they are receiving. Such instruction may be less intensive than the bilingual program at the same school system, be focusing on different curricula objectives, be using materials which emphasize different content, etc. Thus it becomes necessary to control for these programmatic differences, or at least understand their implications, so that plausible interpretations of bilingual program effects, especially unanticipated consequences, are made. Cook and Campbell (1979) have presented various quasi-experimental designs which partition the intensity of treatment and thus permit plausible causal inferences.

The above represents the logic of using comparison groups, i.e., to be able to state with some degree of confidence that trends in follow-up data are more likely associated with being in the program than equivalent students (or for that matter, bilingual teachers relative to comparable interns) could have occurred in their score pattern solely over the passage of time, without being exposed to such instructional benefits. However, other related issues affect the use of comparison groups in follow-up research studies.

One should retain the original comparison groups from the secondary or prior study, and similarly track their progress on the dependent measures of interest. In a planned follow-up study this would be desirable. However, there may be some merit to allocating proportionately less effort to tracking that initial comparison group (even in a post hoc follow-up study), and instead constituting a new comparison group which represents characteristics of greater interest to policymakers. Such a circumstance may arise if time has passed after the conclusion of the secondary study. Shifts in congressional and OBEMLA program management priorities may redefine the value or payoffs of follow-up results if, for example, not only LES/NES Spanish elementary school children were tracked into high school levels, and their bilingual Spanish comparison group grade counterparts, but also a comparison group of matched English dominant children also was added to the follow-up design. In a sense, greater variation in the nature of the three groups now exists, ranging from LES to bilingual to English dominant. However, in this example, the latter group actually was added to address the policymaker questions: Do LES children offered several years of bilingual program instruction tend to "catch up" to their English speaking counterparts at the same grade levels? At what point in their bilingual schooling does "exit" into English speaking classrooms occur? Such questions may assume priority over the types of questions justifying the initial comparison groups (i.e., what is the relative effectiveness of bilingual education schooling for different levels of linguistically dominant children?) Thus, adding a comparison group at the onset of the follow-up time period, while phasing out or tracking the initial comparison group to a lesser extent, is a worthwhile allocation of resources.

The example offered above also illustrates the use of multiple comparison groups. If both comparison groups were maintained, a wider range of group characteristics would exist and a greater range of alternative explanations for data trends could be dealt with, making the addition of this comparison group (i.e., English dominant) a policy-relevant and timely one. Both comparison groups would now be maintained and tracked with respect to patterns of data. Each group and its "baseline" role for interpreting potential bilingual program effects is needed to increase the "research credibility." (Morell, 1979).

As a footnote, the ingenuity of the investigator sometimes is needed to establish meaningful comparison groups. Understanding the ramifications of the study objectives is essential here. Other logical comparison groups for follow-up studies therefore might include:

- For Student Follow-ups - high school graduates of bilingual programs compared with those of the same language group who did not participate in the program and English speaking high school graduates who also were not in the program; comparative studies of those in the program yet who were exposed to different instructional models; or one language group, compared to others also in the program.

- For Teacher Follow-ups - new teachers compared with those of specified levels of experience; those working with teacher aides compared with those who generally do not have aides; bilingual certified teachers compared to those who are not; those with extensive inservice training compared with those having relatively slight amounts of training.

- For Program Follow-ups - programs who have particular instructional models compared with those who use other models; programs with heavy emphasis on parent and community involvement with those with less emphasis on such involvement; newly funded compared to those in the last year(s) of the OBEMA funding cycle; those solely or predominantly funded by Title VII, compared to those proportionately less funded by Title VII, compared to those programs predominantly funded by non-federal sources.

4.5 Effects of Time Prior to Follow-up and Thereafter

Morell (1979) succinctly stated the time-bound issues which affect the internal validity of a follow-up study when he pointed out (p. 241):

> "The longer a follow-up study continues the less it resembles program evaluation, and the more it resembles research on the life styles of particular groups of people. As time goes by, so many factors intrude to influence people's behavior and life styles, that it becomes very difficult to sort out the relationship between people's lives and the effects of a single, identifiable program."

Thus some consideration needs to be given to: (a) how much time has elapsed between baseline and (the onset of) follow-up measurements; (b) how widely spaced is each wave of follow-up measurements; and (c) what is the total span of time (in years, perhaps) which the follow-up data collection period should cover. Furthermore, in a planned follow-up study the investigator may have some control over the time elapsing between baseline and follow-up measurements, whereas a post hoc follow-up study is much less likely to have this latitude (unless an arbitrary decision is made when to begin the data collection for a follow-up study).

Related to these overall time considerations is one of a psychometric nature. A certain, optimal amount of time needs to elapse before program effects and operations will impact on the student or teacher or program. Too short a time span between measurements (baseline and follow-up) will not allow one to detect, or may cause one to underestimate the amount of actual impact or unanticipated consequences (which has policy implications). Too long a time span may produce a "washout" or progressive diminishing of the actually obtained effects due to (a) intervening events in students' life experiences and/or (b) attrition due to mobility of the subject population and comparison group individuals becoming unavailable or unreachable for reinterviewing and retesting. Real difficulties can exist in locating a sample if the between-measurements time span is too great and groups are not maintained. Such attrition generally becomes more pronounced with the passage of time (St. Pierre, 1980). If it interacts with the effectiveness of the bilingual program treatment (e.g., an individualized instruction program feature) such that those who benefitted most leave the measurement cohort to a greater degree than those who did not benefit as much, then the obtained results from those who are contacted may spuriously indicate that the program had a limited effectiveness. Trend lines of treatment and comparison individuals may appear to be fairly similar in their data pattern, although this is not actually the case.

There are differing forms of longitudinal, follow-up designs, each with their relative pros and cons or tradeoffs. Some of these approaches rely on matched longitudinal cohorts (the term used by Hilton and Patrick, 1970) in which the same group of students or population of students (e.g., Native American graduates of elementary school bilingual programs) are tracked and repeatedly tested, interviewed, or sent mail questionnaires at different times. Other designs use unmatched longitudinal cohorts (also defined by Hilton and Patrick, 1970) in which intact student groups, i.e., all the students tested each time, are used. As Hilton and Patrick (1970, p. 15) put it, "In other words, when the sample statistics are computed they are based on all the students who were present in the school at the time of testing whether or not they were in the original sample."

This might represent all fifth graders being served in the second year of program operation, although they were not necessarily part of the initial student body served in the first year of funded operation. Rather, the tested group may include those who moved into the school system in the second year, those who are being served because of an attendance zone realignment between schools, those who are being served because of an instructional activity clustering reorganization across targeted schools, etc. In another example, each year's teaching staff is not the same as in the previous year. To collect data on the unmatched cohort is apt to bring in data on individuals with different experiential characteristics, philosophies of bilingual education, etc.

Thus, there are the simple longitudinal designs and extended longitudinal designs (presented by Labouvie, in Bentler, Lettieri, and Austin, 1976). The simpler designs deal with individuals sampled from some target population and measured repeatedly on two or more occasions, with internal and external validity being a function of how well various sources of error are controlled. The extended longitudinal designs include:

- Time-sequential designs, which do not provide longitudinal observations of intraindividual changes, but which do provide information about general cultural trends as a background against which to evaluate the impact of specific intervention programs; this approach is confounded by cohort effects, i.e., groups born at different times;
- Cohort-sequential designs, which use a set of cohorts being observed at different age levels, providing a longitudinal series for each of several generations, an approach requiring an inordinate amount of time for completion (based on the range of cohorts chosen); and
- Cross-sequential designs, which use a fixed set of cohorts being observed on several occasions using repeated or independent observations (i.e., different topics or variables are dealt with); this design has the greatest practicality of the three approaches, and possibly might be feasible for the present study.

To illustrate what the three designs look like schematically, Exhibit 5 presents each (from Labouvie, in Bentler *et al*., 1976, pp. 51-52).

EXHIBIT 5

FORM OF THE TIME-SEQUENTIAL, COHORT-SEQUENTIAL, AND CROSS-SEQUENTIAL LONGITUDINAL DESIGNS

(a) Time-sequential Design: Four Ages are Measured at Four Times

Ages	Time of Measurement			
	1972	1975	1978	1981
8	1964	1967	1970	1973
11	1961	1964	1967	1970
14	1958	1961	1964	1967
17	1955	1958	1961	1964

Cohort/Time of Birth

(b) Cohort-sequential Design: Four Cohorts are Measured at Four Ages

Cohort	Ages			
	8	11	14	17
1965	1973	1976	1979	1982
1962	1970	1973	1976	1979
1959	1967	1970	1973	1976
1956	1964	1967	1970	1973

Time of Measurement

(c) Cross-sequential Design: Four Cohorts are Measured at Four Times

Cohort	Time of Measurement			
	1972	1975	1978	1981
1957	15	18	21	24
1960	12	15	18	21
1963	9	12	15	18
1966	6	9	12	15

Ages

Source: Labouvie in Bentler, Lettieri and Austin, 1976.

However, any of these designs needs to recognize the needs of those who funded the study and the overall purposes of the study, so that the design is as sound a "fit" as possible to those agencies and their needs. Thus questions need to be addressed in selecting the type of follow-up study such as:

- How long a follow-up study is feasible, in terms of the number of data collection points or the number of school years?
- How can the follow-up study design best take note of ongoing changes in the nature of: the local bilingual program, any of its significant instructional features, local contexts or even federal-wide program-related changes (such as changes in funding, regulations and types of programs being funded), so that the funding agency maximally benefits in terms of quality, up-to-date planning information?
- What are meaningful comparison groups, and how many should be used?
- How wide a time span between measurement points should occur, to optimally detect program effects?
- Should certain measurements on certain variables be gathered on particular follow-up occasions, and due to the nature of the dimension being measured (e.g., long term attitude change), gather that information at later stages in the follow-up data collection period?

All of these issues have conceptual, interpretative, policy-relevant, program planning design, and statistical implications, and therefore need to be addressed. Also, if policy and program representatives are aware of the implications and tradeoffs of proceeding in one design direction rather than another, they will be better able to use the results gathered at the time that they need it. Such an approach is currently being used with multiyear evaluations which schedule data collection and the number of such occasions or waves so that they can become inputs into congressional reauthorization hearings.

4.6 Cross-Sectional Follow-up Designs

It may be that an alternative to longitudinal follow-up designs is needed, despite all the benefits cited for them by a number of evaluation authorities, among them Morell (1979), Tucker (1979) and St. Pierre (1980). A cross-sectional design alternative to the longitudinal approach may be needed. This consists of gathering data at one point in time from students of different ages or different grade levels, thus collapsing the time dimension which would have allowed, for example, sixth graders to be successively tested at each year throughout their middle/junior high and high school experiences. However, on balance, cross-sectional designs have their own relative merits and deficiencies, summed up by Broedling, Wiskoff and Herbert (1975, p. v) as follows:

> Longitudinal designs are superior to cross-sectional designs... because the former make possible the identification of historical or maturational factors, allow for a clearer specification of causal relationships between attitudes and behavior, eliminate between-group error variance, require a smaller number of subjects, and are more applicable to studies where experimental intervention is required.

More specifically, Hilton and Patrick (1970) created a "balance sheet" appraisal of cross-sectional designs compared to either matched or unmatched longitudinal designs. Exhibit 6 contains that comparison.

EXHIBIT 6

SUMMARY OF SOURCES OF DIFFERENCE IN THREE TYPES OF DATA WHEN INTACT SCHOOL GROUPS ARE TESTED

	Cross-Sectional	Longitudinal	
		Matched	Unmatched
Age Difference	x	x	x
Cohort Difference	x		
Time Difference		x	x
Equating Errors	x	x	x
Retest Effect		x	x
Cohort Changes	x		x
Selection Effect		x	

Source: Hilton and Patrick, 1970.

From examining Exhibit 6 from Hilton and Patrick (1970), it should be clear that some of the issues in constructing a sound longitudinal research plan are methodological while others are "real-life." The methodological issues quite certainly need to be dealt with, because they have strong ramifications for the generalizability of results, since they include such factors as the following, which are not present to the same degree in cross-sectional research (as pointed out by Hilton and Patrick, and Broedling et al.):

- Retest effects - extensive testing brings along its own shift in scores, due to the same individuals or programs becoming "sensitized" to the fact that they are in the study.

- Age change - maturation effects, as the sample is now older than when originally tested or data gathered on them (even from behavioral sources).

- Time change - an environmental change in which an event, such as changes in entry-exit requirements, influenced all members of the sample regardless of age.

- Equating errors - spurious change introduced by one sample being given one test and an alternative form being given the other, comparison group sample (correspondingly, this could apply to different instruments or data collection procedures being used for each sample).

- Selection effects - this refers to non-random loss of cases when a matched longitudinal cohort approach is desired; the reasons for attrition are related to the topic being measured.

- Cohort change effects - the successive samples studied are not randomly sampled from a stable population, e.g., the intact groups (100% of each group or a random sampling) of all the students in targeted grades available at the time of data collection. Between such stages the cohort will change, based on mobility, transfers in, etc. These shifts will result in net changes in the composition of the cohort which will result in mean score changes -- changes which are not age changes or time changes. This corresponds to an unmatched longitudinal cohort design.

In general, cross-sectional data by its very nature precludes a time difference, but because it includes a different group of individuals at each age, grade level, length of time in the program, etc., a cohort difference is involved. Furthermore, on balance, Morell (1979) stated that "all efforts should be made at increasing the internal validity of data even at the expense of decreased generalizability."

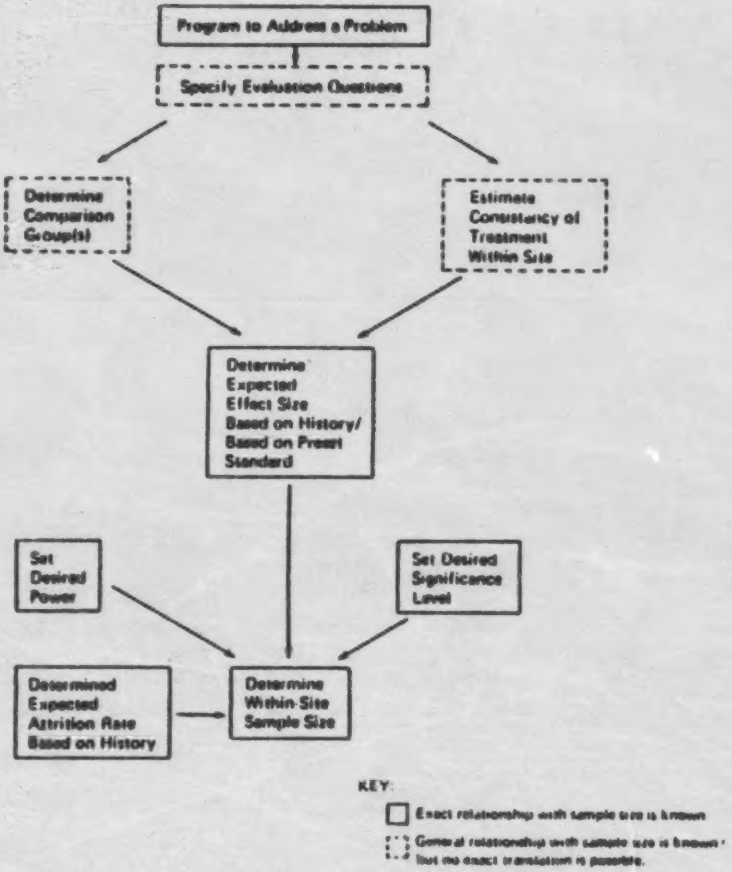
4.7 The Problem of Attrition

There are many ways in which the number of individuals or programs originally included in a study can substantially decline, to the point that inferences and conclusions drawn from the study findings are seriously compromised. For the most part, these attrition processes are not random, and it is difficult for researchers to convince themselves that random attrition has occurred.

The attrition rate can be a sizeable one, and should not be ignored. In fact, St. Pierre (1980) suggested that a determination of the expected attrition rate based on that (or comparable) program's history be directly built into sample size decisions (see Exhibit 7), along with considerations of effect size, significance levels being used and desired power levels. St. Pierre reviewed similar longitudinal evaluations and concluded that attrition rates of about 50% over time might occur. For example, Project Developmental Continuity (PDC) had a 60% attrition rate over a five-year time period. Thus, St. Pierre suggested that having about 120 treatment and 120 comparison group students available per site at the onset of a program would provide for relatively large effects being detected, at a liberal significance level and a moderate level of power. However, this step would exclude relatively small programs. Since 1979-80 ESEA Title VII grantees had a median number of 324 students served per program, the St. Pierre suggestion corresponds to a 37% sampling rate of all children in such programs, and a much higher sampling rate at even smaller programs. Such a rate may be plausible but administratively disruptive. To fit bilingual programs, certain assumptions made in the article, although they are sound ones, may need to be relaxed. Thus, those involving effect size can perhaps be lowered, and smaller sample sizes per site used in a follow-up study.

EXHIBIT 7

PROCESS BY WHICH NECESSARY SAMPLE SIZE MAY BE DETERMINED


Program to Address a Problem
Specify Evaluation Questions
Determine Comparison Group(s)
Estimate Consistency of Treatment Within Site
Determine Expected Effect Size Based on History/ Based on Preset Standard
Set Desired Power
Set Desired Significance Level
Determined Expected Attrition Rate Based on History
Determine Within-Site Sample Size
KEY:
Exact relationship with sample size is known
General relationship with sample size is known, but no exact translation is possible.

Source: St. Pierre, 1980.

The types of circumstances in which attrition can occur include the following (primarily taken from Marshall's 1976 examination of Project Follow Through attrition processes):

- Subject attrition - the number of individuals included in the evaluation at its onset diminishes with the number of data collection occasions, due to mobility, absenteeism, refusals to participate, etc.
- Item attrition (or omission) - certain sections or items of questionnaire instruments are left uncompleted on one or more of the data collection occasions, having the effect of different students having completed different sections or items of the instrument, making computed between-item correlations somewhat unrepresentative of the full number of students in a study (an attrition pattern generally related to either the content or format of the item(s) involved).
- Data collection attrition - individuals were excluded because they were not contacted by the data gatherer, or because the contact did not yield the information required for analysis.
- Data processing attrition - individuals included in the design failed to be included in the study data records.
- Analysis attrition - individuals included in the evaluation design were deemed invalid for the analytic design.
- Design attrition - programs (or individuals) initially involved in the study were excluded because of changes in study focus or rationale (which in the case of the Head Start Longitudinal Study occurred and removed the St. Louis site for administrative reasons).
- Program attrition - in which children changed treatments or changed models.

Any and all of the attrition processes just mentioned are likely to have the same two general effects, i.e., of (1) "changing an ordinarily simple statistical analysis into a complex one" (Kim and Curry, 1978), and (2) biasing study analyses and weakening their credibility to policymakers, since certain -- perhaps policy-relevant -- segments of the study population are excluded from the study. The first two types of attrition processes are discussed in this section because they are more central to the overall goals of follow-up studies. The other types of attrition processes generally represent administrative, data management or cross-over design confounding influences, and as such, are beyond the scope of this paper.

Thus, the remainder of this section will deal with some of the patterns of subject and item attrition which occur, and the alternative approaches which have been offered in the literature -- with varying degrees of success -- for dealing with such attrition.

Some attention needs to be paid to the presence of attrition effects. On some occasions attrition can be a random process and therefore it does not overly bias the study findings. However, there are certain theoretical issues that need to be considered in detecting and dealing with the presence of biasing attrition, viz.:

- What are the reasons offered by subjects or associated with subjects for leaving the study data base?
- Are the factors associated with the attrition rate also related to the pattern of data on outcome/impact measures?
- Are the attrition rates site-related, such that differential attrition rates exist for particular bilingual programs or subsets of data collection units?

With plausible answers to these questions it then would be possible to estimate the magnitude and direction of bias, and then turn to the most useful approach for dealing with attrition effects. To address these issues Bean (1979) suggested a path analysis model to explore the determinants of attrition. Cook and Campbell (1979) also suggested that these issues be explored sequentially. That is, first one should test for the attrition rate being equal in all groups, then determine if the reasons for dropout are similar for both the treatment and comparison groups, then determine how equivalent or comparable the data collection units remaining in the study are on the background variables measured before the study began, and finally, turn to issues of assessing the equivalence of pretest (i.e., baseline) measures. The latter represent the best available single estimates of bias.

Hypotheses should be used to explore the nature of differential attrition as advocated by Cook and Campbell (1979). Such hypotheses developed in the context of bilingual program issues, and any tertiary data gathered, will provide a range of estimates of the effect of attrition phenomenon. As

Cook and Campbell (1979) stated: (p. 364) "...the estimates will bracket a range of contradictory outcomes and will thus serve as a warning that attrition artifacts may be masquerading as treatment effects." One such hypothesis which they offered and which could apply to bilingual program settings, was that "the children in educational experiments who persist in voluntary exposure to an educational innovation will be the more able, and the less able will drop out in greater numbers." In general, Cook and Campbell recommended that a "concentrated effort" be made at the onset of a study to clearly understand all of the reasons why comparison group members (termed "control group members" in their discussion) might differentially leave a study. It might even be able to reduce some of the attrition (program, parent, student, etc.) with advance information on study purposes.

Some of the adjustments to compensate for missing data include the following:

- Only data from students present at all testings can be used (a selective bias in itself which is likely to weaken the power of the analysis);
- Cases most similar to the missing cases can be weighted more heavily in the analysis than those least similar;
- Methods suggested by Reichardt (in Cook and Campbell, 1979) can be used to suggest treatment effects in a non-equivalent control group design;
- Dummy variables can be put into the model (indicating missing data on a particular measure) to adjust treatment effect estimates for the fact that the attrition may be correlated with the outcome; and
- Missing scores based on data from students in the same group can be imputed, using cell means or medians, or by iteratively (or non-iteratively) regressing missing scores on the vector of available data.

The field has mixed opinions concerning how best to adjust for systematic pre-existing differences arising from attrition of data units. Some authorities suggest (sometimes insist) that no adjustment is perfect, and that any adjustment performs inexactly, never capturing all of the relevant differences. Other authorities say that programmatic adjustment, though not perfect, still is useful, and thus, should be employed carefully. Since even indications of impact could be useful to know (if they are present), adjustments to understand the magnitude of differences due to pre-existing differences or attrition might be used to provide policymakers with credible impact and trend information.

There are similar issues which exist in dealing with item attrition or missing data. The researcher needs to know when and under what conditions the problem of missing data can be considered trivial. Kim and Curry (1978) also pointed out that "the smaller the relative proportion of missing information, the larger the sample, and the more random the missing information, the less troublesome the missing data problems." However, what proportion of missing information should be considered relatively small? Kim and Curry categorized the patterns of missing item information as falling into five groups:

- It was randomly produced;
- Missing data on a given variable were dependent on the value of another variable;
- Certain underlying values such as education or cognitive ability determined whether information is missing or not (since with the former variables, conditional estimates of the missing information can be made);
- Whether information is missing on a given variable was dependent on the levels of the variables itself (as when respondents are reluctant to reveal high levels of family income); and
- Missing data was a product of a combination of two or more variables (as when those with high earnings are unwilling to reveal their low educational levels) making estimates of the missing data from other existing data rather misleading.

Kim and Curry (1979) reviewed the relative merits of listwise deletion approaches to missing data, i.e., only include those subjects with complete data on the variables of interest and pairwise deletion, i.e., include those subjects having data on the pairs of variables included in a particular bivariate correlation coefficient. The effect of the latter approach, when excessive amounts of missing data occur, is to produce results using certain variables which are non-applicable to the full number of subjects in the study.

Kim and Curry also review the rationale for estimating missing values (rather than the deletion approaches mentioned directly above). They point out that using estimates of missing information can serve one of three main purposes: (1) it would simplify calculation of statistics (a reason outdated by computers and programmed least-squares algorithms), (2) it would improve parameter estimates and (3) it would retain as many cases as possible in constructing scales out of many variables. Both of these latter two reasons are plausible; they serve the overall purpose of generalizing from sample data, and especially from a complete data set with no missing data. However, for the most part, various measurement properties of survey (and questionnaire) data tend to make any estimation process to recover missing data a complicated one. Convenience and feasibility also affect whether listwise deletion, pairwise deletion, regression approaches or other techniques should be used.

Furthermore, a sample size of over 1,000 may mean that the choice of method is not a practical consideration. Alternative techniques would work about equally well, since interitem intercorrelations of existing data would be stable, etc. However, if the amount of missing data were relatively slight, yet confined to particular bilingual programs with instructional models having a great deal of interest to policymakers, then an estimation approach which "recovers" the missing data by inserting likely values would be much more useful to users of the study findings than by simply excluding its data (the listwise deletion approach).

4.8 Provisions for Identifying Follow-up Samples

Acknowledging that attrition of subjects from the follow-up study phase is generally undesirable, since various forms of data bias may confound interpretations and conclusions, the question logically arises -- what can be done to minimize such attrition?

Cochran and Cox emphatically answer (1957, quoted in Morell, 1979): "the only complete solution to the problem of missing data is to not have any."

However,certain approaches for dealing with attrition come to mind, and have been used effectively in other evaluation or longitudinal survey research contexts. The best practice is to identify the follow-up sample before or during the original study. This can be done by directly asking potential follow-up respondents, "Where do you expect to be a year from now? Could you supply us with the names of two relatives or friends who could help locate you at that time?" Such an approach was effectively used in a study conducted by Development Associates (Otal, Goldsamt, Wilson, Hopstock, Copping and Toledo (1980). Households were contacted during the first data collection phase, an area-wide administration of a housing and demographic survey instrument. The second wave, one month later, using a socio-economic survey instrument, only included households earlier identified as Hispanic. Having the telephone numbers of these households and information on when persons generally were at home (weekdays or weekends, day or evening) greatly helped to increase the recontact rate, and to reschedule the Hispanic households as needed. Such information was generally easy to obtain if asked for at the conclusion of the interview, since interviewer rapport already had been established as a result of participating in the full interview.

A similar, also highly effective approach was incorporated into Base Year questionnaires used with the National Longitudinal Study of the High School Class of 1972. This study, termed NLS 72, was funded by the National Center for Educational Statistics and was conducted by Research Triangle Institute (Collins and Bailey, 1975; NCES, 1979). Included in the Base-Year student questionnaire was a request for not only the student's name, address and social security number, but such additional data as the name and address of two people who would always know where to get in touch with the respondent, driver's license number and the state, and the name of the postsecondary educational institutions they planned to attend. A similar set of tracing information was requested in Section F of the First Follow-Up questionnaire. That contained, on the inside back cover page, detailed tracing information such as parent and friend addresses and telephone numbers, date of birth, social security number, and driver's license number. Furthermore, individuals who had not participated while in high school, but who now were included, were supplied a return postcard requesting information about the student's current address at that time and anticipated address as of October 1973.

There are other approaches for monitoring the whereabouts of individuals for follow-up study purposes. The "cohort maintenance" activities can develop provisions for tracking students through a school system, perhaps building on any similar, and possibly computerized forms and procedures already being used by any large school systems included in the initial or follow-up study. Such an approach was developed by Development Associates for use in its five-year longitudinal study of the Experimental Schools Program's effectiveness in the San Antonio, Texas, Edgewood Independent School District, a system which is 90% Mexican-American. (See Jones and Davis, with contributions by Theimer, Cervantes, Goldsamt, Toscano and Bernal, 1977).

Related to this is the consideration of how to best organize, in a research data management sense, the data storage and retrieval system, so that all the data collection units retained in the follow-up study phases contribute data which are readily usable for analytic purposes. Such a retrieval system will benefit a great deal by being designed in advance, to fully mesh with the study's descriptive, analytic and policy maker informational needs. If the system also can readily use any identifying and data-retrieval features of a school system's centralized student tracking system, it may become that much more effective. That is, investigators need to recognize the utility of having a research data management system geared to deal with large numbers of students, and sites, and perhaps data collection occasions. Such a system probably needs both computerized and non-computerized features, to handle raw documents, computer tapes and disks, correspondence to sites, follow-up contacts to students, teachers, and administrators, etc.

Such considerations are necessary to achieve the utility of such studies as NLS 72, which dealt with overall post-high school education and work experinces, graduate school and military experiences, college graduate dropout comparisons, educational benefits in the work world for women and minorities, and post-secondary educational benefits of students after they left school. Its data base is so extensive, due to an overall cohort maintenance rate of 93% (and a very considerable effort to achieve it) that approximately 250 research studies have been generated from this data base over the past eight years. (NCES, 1979).

A similar hope is held out for NLS 80 which will focus on high school and beyond, augmenting the data base to include high school sophomores and high school seniors. The study goals are to focus on the impact of major federal programs on blacks, Hispanics and other groups; examine the nature of high school dropout *per se*; and determine the financial plans parents have for their children's post-secondary education. The sample sizes, already quite large, also will be augmented to permit state-level analyses of programs and services on the performance, future education and careers of their own students. (NCES, 1979).

Thus in both of these major substudies, which have substantial follow-up components (NLS 72 is in its fourth follow-up cycle), extensive cohort maintenance, sample retention activities and a comprehensive research data management system are givens. Without these requisite features of the overall study design, an extremely murky picture of educational benefits and experiences would result. The nature and scope of such cohort maintenance activities is outlined in the following section.

4.9 Techniques for Maintaining Follow-up Samples

Attrition is typical in longitudinal studies and it can become very difficult and resource-consuming to control. It can be done, however, given the "right" population, investigator ingenuity, financial resources, and topic issues being investigated (Morell, 1979). This has been aptly demonstrated by: (a) NLS 72 over a four phase follow-up period (Collins and Bailey, 1975), and (b) the intensive efforts of the large-scale test validation study conducted by the U.S. Department of Labor (DOL) in which individuals were followed-up two years after high school graduation, and seven years after high school graduation, each time with highly effective rates of cohort maintenance of the original individuals (Droege, 1971). Thus the techniques indicated in this section mainly come from the NLS 72 and DOL presentations of the approaches used for tracing and retaining cohorts. The discussion is supplemented where useful by material found in Morell (1979) who concluded (p. 232):

> "An infinite number of strategies are available and the correct choice is highly situation specific....the best solution to attrition is to exert all efforts at avoiding the problem. Research shows that surprisingly large numbers of respondents can be located, that various techniques for eliciting cooperation do work, and that the missing data problem can in fact be greatly diminished."

To begin with, Morell reviewed the literature and found that any one of four tactics can effectively reduce attrition:

- Degree to which the cover letter is personalized, and use of first class mailing to increase rates of retention.
- Explaining the objectives of the study to participants increases retention rates.
- Repeated reminders to nonrespondents will be useful in reducing their number.
- Payments to respondents for their participation may tend to maintain the cohort, although the possibility of differential bias may arise and therefore needs to be considered.

With respect to the cover letter and any other written communication to members of the bilingual program follow-up samples, we might extend this approach and have letters, newsletters, brochures, etc. written in the L1 language. Using bilingual versions of written materials may increase or at least hold constant the response rate. Such an approach was effectively used by Development Associates in the series of interviews conducted with Hispanic households in the Adams Morgan - Mt. Pleasant section of Washington, D.C. (Otal and Goldsamt, et al., 1980). Survey instruments were both in Spanish and English, administered by local bilingual residents, and were supported by bilingual: "thank-you-for-participating" letters, survey goal brochures, and rescheduling "sorry-we-missed-you" notes slipped under residence doors. The cumulative effect, spread both officially and through word-of-mouth, was to convey the impression that a culturally sensitive organization was collecting data for the good of the community.

As indicated by the above approaches, no single technique is likely to retain individuals in the sample cohorts. Rather, an organized, coordinated and complementary series of approaches is needed. Exhibit 8 portrays the wide range of tracing efforts which were used by the NLS 72 study (Collins and Bailey, 1975) to obtain an overall response rate of 93% of respondents to the First Follow-up Survey, from a total questionnaire mail-out of 23,020 forms. To achieve this level required a series of activities in a concerted flow. The activities which were actually performed are so clearly described by Collins and Bailey (1975, pps. 3-8) that they are reproduced here in only slightly adapted form.

> "The survey plan for First Follow-Up involved from the start an unusually extensive effort both to enlist the support and cooperation of potential NLS respondents and to obtain updated addresses of this highly mobile group. Two quite-similar newsletters were prepared early in July 1973 with the purposes of relaying information just coming out of Base-Year data analysis and reinforcing the goals and purposes of NLS communicated during Base-Year data collection in the high schools...The "resurvey" newsletter required an additional short explanation covering how and why RTI and NCES obtained the sample member's name...
>
> The first step in data collection for First Follow-Up involved an extensive tracing operation to update name and address files. Research Triangle Institute received from the Base-Year contractor (Educational Testing Service) a name and address file...
>
> Prior to the mailout of First Follow-Up Questionnaires 102 individuals were deleted from the list of eligibles for the survey due to duplicate names, incomplete addresses, deceased, etc...Newsletters with accompanying return postcards were developed and mailed in July and September 1973 to the augmented file of 23,122 sample members to encourage participation and to use as a vehicle for updating names and addresses. Postcard returns were used to update records; and when newsletters were returned by the Postal Service as undeliverable, telephone tracing procedures were used to obtain current addresses where possible...
>
> RTI established a Telephone Tracing Department to handle NLS tracing activities, and trained staff were recruited, special equipment installed, and a detailed manual was written...
>
> Tracing began early in August 1973 as the Postal Service returned undeliverable mail from the newsletter mailouts, and operations continued through January 1974 to include tracing for questionnaires which were returned as undeliverable...

> The Tracing Department successfully traced over 90 percent of the more than 2,000 cases it worked on. Although almost all of the tracing activities were conducted by telephone, a large portion of the 177 unsuccessful tracing cases included mail as well as telephone attempts to obtain a current address. Mail was used in those instances where students, parents, and other contacts did not have a telephone, and also in an effort to obtain current addresses for some colleges and state Motor Vehicle Departments... Fourteen of the cases that were not successfully traced were sent to a local credit bureau for processing. The credit bureau was unable to obtain a current address on any of these cases...
>
> As new addresses were obtained, whether from tracing activities, returned mail, or other sources, the computer file of names and addresses was updated and, when necessary, an additional mailing of a questionnaire was made. Active mail return efforts continued through December 1973; and by early January 1974, the questionnaire return rate by mail was over 60 percent. This was followed by a planned sequence of reminder postcards, additional questionnaire mailings, and reminder mailgrams to nonrespondents---all controlled and updated by a computerized automated survey support system...
>
> In addition to these materials, a letter from NCES accompanied the second questionnaire mailout to nonrespondents, a letter of thanks from RTI was sent to all mail respondents, and a computer-generated self-mailing Identification and Change of Address Card was sent to all First Follow-Up participants after survey operations were completed...
>
> The names and available tracing information for over 8,000 sample members who failed to return their questionnaires by mail were then turned over to the Bureau of the Census for personal interview in accordance with a Bureau contract with the U.S. Office of Education. This personal interview phase of first follow-up data collection continued through March 1974, at which time the overall response rate was increased to about 93 percent."

In much the same way, Droege (1971) reported on three years of high schools' graduating classes (a total of 168) being asked to complete an "address sheet" in the spring of their senior year. This sheet asked for present address and phone number, social security number (each student was assigned one if he/she did not have one already), and names, addresses and phone numbers of others likely to know the location of the individual two years later. At that time employment service personnel in 19 participating states attempted to recontact the individual, to complete a two-page 28-item information form providing job and education data. The seven-year follow-up was termed "considerably more difficult" since it entailed completing a detailed questionnaire.

EXHIBIT 8

TRACING SOURCES OF INFORMATION

1. Name, address, and telephone number of parents, guardians or other relatives of the individual.
2. Names, addresses, and telephone numbers (furnished by the individual and other contacts) of the two people who would always know how to get in touch with him.
3. Name and location of the postsecondary school the same member has attended or planned to attend.
4. Name and location of the sample member's employer, as reported in the first follow-up questionnaire.
5. Neighbors of the sample member or his parents.
6. Principal or other contact at the secondary school attended.
7. State or registration and identification number of driver's license.
8. Local government agencies.
9. Armed forces locator services (or rosters as supplied by DOD for the first follow-up).
10. Institutional records (i.e., prison, police, mental institutions, etc.)
11. Local credit bureau or other similar organization.

Source: Collins and Bailey, 1975.

Of interest is that the national office of the relevant DOL office provided general guidelines for doing follow-ups, but state employment services were encouraged to develop the specifics of the approach to be used and techniques to be applied. There is an analogue to this for bilingual education programs. It would appear to be very useful to meet with local program personnel and community representatives to learn what approaches work best in each setting. Perhaps the "informant network" is better developed in certain local contexts than others, while town meetings may be more suitable for school parents in other settings. Matching the goals of the follow-up cohort retention approach to the local setting would seem to be highly desirable.

The overall effect of the state DOL office tracing and recontact activities was to achieve a response rate of 93%. This was not a static rate. Those doing the follow-up actually improved the effectiveness of their two year follow-up approaches, going from 86% for the class of 1959, to 94% for the class of 1960, to 99% for the class of 1961. As Droege pointed out (1971, p. 28), "This record is an indication that it is not unreasonable to set goals of close to 100% success in follow-ups that do not require detailed information that must be supplied by the individual himself."

Taking this as a lesson, perhaps the scope of data collection effort for bilingual education follow-up studies should be seriously considered, to assess how much of a respondent burden actually is involved. Such a concern meshes very well with Morell's 1979 desire to maximize internal validity at the expense of generalizability. Certainly, a minimal attrition rate is a step in that direction.

Droege's presentation of DOL recontact efforts after seven years also indicates a fairly successful level of cohort maintenance, 84%,albeit less than the two year rate. This is an indication of the greater difficulty in obtaining more detailed information after a longer waiting period. In fact, whereas it took about 2.03 total contacts to obtain a successful contact with two year follow-ups, it required 3.29 contacts to obtain a successful contact after five more years with the same individuals (i.e., a total of seven years). Thus 62% greater effort was expended in the seven-year follow-up activities.

Exhibit 9 indicates what techniques actually were used. The best techniques in terms of percentages of attempts that were successful were personal visits, telephone calls, information from local employment service offices, and individuals in the community. The most useful techniques were A, B, G and I (listed in Exhibit 9), providing 96% of the two-year successes, and 85% of the seven-year successes. The most used technique in both follow-ups was a letter to the individual using various sources for address and mailing up to five letters to the same address. Although the technique declined with the number of times the letter was sent to the same address (as would be expected), even third and fourth letters resulted in a substantial enough return to warrant their use, rather than resorting to more expensive techniques immediately (see Exhibit 10).

Of particular interest here is that individual states were more successful than others, a factor which was strongly related to the number of techniques used. Thus for 1960 graduates, the nine states that obtained information for over 99% of their potential samples used about nine techniques, compared to about five techniques used by the other 10 participating states that had less than 99% of their potential samples. Similar findings were obtained for seven-year follow-up efforts.

This has implications for conducting follow-ups in the field of bilingual education. It suggests that a variety of techniques is likely to account for a high percentage of success, but also that certain states may be more effective than others. Considering also the fact that differential sources of attrition may exist at certain bilingual sites and not at others, this suggests that perhaps states with greater or more comprehensive bilingual education priorities, visibility and the presence of a state plan may, in conjunction with effective tracing techniques, be more able to maintain their respective follow-up samples than other states. At the local level, this may suggest that a more visible bilingual project director, perhaps co-signing a cover letter being mailed with a follow-up questionnaire, may induce higher response rates and cohort maintenance. These ideas are suggested here,since in the Droege study, states were allowed to adapt the tracing techniques as needed. Simply using more techniques than others may have not been the full

EXHIBIT 9

OVERALL ANALYSIS OF EFFECTIVENESS OF TECHNIQUES

Techniques for individuals in potential sample	2-yr. follow-up attempts to obtain the information			7-yr. follow-up attempts to obtain the information		
	Total *N*	No. successful	%	Total *N*	No. successful	%
A. Letters to individuals in sample using address shown on address sheet	31,610	16,030	50.7	33,622	10,070	30.0
B. Letters to individuals in sample using address obtained from others on address sheet	3,220	769	23.9	5,157	1,396	27.1
C. Letters to individuals in sample using address obtained from city or telephone directory	209	72	34.4	1,637	520	31.8
D. Letters to individuals in sample using address obtained from Department of Motor Vehicles	155	44	28.4	1,393	352	25.3
E. Letters to individuals in sample using address obtained from high school attended	420	175	41.7	1,445	331	22.9
F. Letters to individuals in sample using address obtained from other sources	13	4	30.8	1,450	436	30.1
G. Personal visits	295	256	87	1,463	739	51
H. State agency "wage item" file	283	143	51	277	80	29
I. Telephone calls	1,333	893	67	1,589	630	40
J. Local office records and employee knowledge	178	114	64	170	78	46
K. Information obtained from individuals in community	88	60	68	382	90	24
L. Other techniques						
High school counselor	107	93	87	79	49	62
Letter to parent	318	138	43	652	162	25
Local post office	0	0	0	67	23	34
Miscellaneous	5	5	100	330	144	44
Total	38,234	18,796		49,713	15,100	

Source: Droege, 1971.

EXHIBIT 10

ANALYSIS OF EFFECTIVENESS OF LETTERS TO INDIVIDUALS IN THE POTENTIAL SAMPLE

Mailing	2-yr. follow-up attempts to obtain the information			7-yr. follow-up attempts to obtain the information		
	Total *N*	No. successful	%	Total *N*	No. successful	%
1st letter	22,582	11,943	52.9	23,887	8,136	34.1
2nd letter	9,567	4,075	42.6	12,613	3,368	26.7
3rd letter	3,342	1,041	31.1	6,761	1,360	20.1
4th letter (or more)	136	35	25.7	1,443	241	16.7

Source: Droege, 1971.

reason for the effectiveness of these states. The form and content of the communication with respondents may have made a difference; these perhaps are adaptable to bilingual settings.

4.10 Related Cost Considerations

A large number of tracing techniques used over a period of time, by a set of individuals having prime responsibilities for conducting these activities, necessarily requires study resources. Increasing postage rates, even with bulk mailing permits and word processor or computer generated "personalized" cover letters also commit study resources. Most bilingual education agencies lack such a level of resources. Thus the issue of how much does it cost to trace and maintain cohorts of samples is a real one.

Most studies, even the Project Talent follow-up study of 88,000 participants, did not provide a complete breakdown of all costs for the entire study. However, Morell (1979) reported that other investigators had estimated that the overall cost per respondent sought (for questionnaire production and mailing costs) was 17.5 cents, and that the cost per questionnaire received was 58.6 cents. This cost figure was based on four mailings which resulted in a total of 61,300 returns, or a 69.7% response rate.

Panos and Rice (1967) performed an analysis of the costs associated with different mail follow-up techniques to reach hard-core nonrespondents who were entering freshmen at a national sample of regionally accredited four-year colleges and universities. By selecting a random subsample of the 23,673 hard-core non-respondents, a set of 665 subjects resulted who were contacted with different combinations of:

- type of cover letter (autotyped with a personalized greeting and signature vs. mimeographed with impersonal greeting and signature);
- type of outgoing postage ("live" stamp vs. metered stamp);
- class of outgoing mail (certified vs. special delivery vs. first class).

The costs per treatment included: postal charges, materials, special handling (as for autotyping and personalized signature) and clerical costs at the rate of $1.25 per hour.

Findings showed that the cost per respondent varied from $1.60 to $3.38. The cost per subject contacted by using a mimeographed-metered-stamp first class approach was 70 cents. However the cost per return (probably the key factor) was $3.16, since a response rate of only 18% resulted. This rate was the least successful of any other combination in the 12 treatment cells defined by combinations of the three factors. This was a surprising study finding. Furthermore, and paradoxically, the cheapest method in terms of cost per respondent also tended to be the best in terms of rate of response. This was the combination of using a live stamp, personal letter (autotyped) and special delivery,which had a 79.4% response rate and cost only $1.60 per respondent.

Panos and Rice concluded that cost levels will in part be determined by the purposes of the non-respondent follow-up. If the overall goal is to obtain a minimal response level, then several optimal ways of allocating the funds available are suggested by the findings. If the purpose of the follow-up of non-respondents is to increase the percentage of respondents, an initially more costly method should be used. If the main concern is to estimate the nature of bias among respondents, a less costly technique should be used.

On balance, Panos and Rice concluded that the initial wave of questionnaires should be mailed out using a relatively less expensive mail technique. (For example, the autotyped personal letter-live stamp-first class combination had a cost of only 88 cents per respondent contacted.) Then, investigators should select a method to ensure a high response rate among non-respondents in the follow-up contact wave, so that a more stable estimate of bias among early respondents can be made, and the overall rate of return can be increased.

As can be seen from the preceding discussions, response rates vary with the extent and type of follow-up technique, the number of years between contact, etc. The methods can be effective but potentially costly and time-consuming, although they benefit from staff personnel being assigned to conducting the tracing procedures. This can be a full-time task.

Such efforts may be somewhat eased if existing data retrieval systems can be used wherever possible. Such systems might include: those developed for the NLS 80 study which will include Hispanics; the computerized system being developed for the selection of bilingual programs, program directors, teachers, and teacher aides for the Classroom Instructional Component of the ESEA Title VII Program study; Project Talent updated address and geographic location files; etc. Using such systems wherever possible may tend to make the tracing procedures relatively more cost-effective.

If these approaches are not fully feasible, then it becomes necessary to rely upon the ingenuity of the investigators, knowledge of the local contexts and bilingual programs involved, and any "detective" or traditionally accepted survey research approaches which can be adapted to current evaluation needs. In this regard we have mentioned bilingual letters and materials to motivate respondents, and using bilingual telephone interviewers to re-contact respondents to mail back survey forms. Another approach would be a joint signing (perhaps on government agency letterhead) of the study director and the primary user of the follow-up study findings, e.g., the Director of OBEMLA or the Part C Coordinating Committee.

In the final analysis, some sensitivity is needed to assess how the particular respondent population can best be traced and retained within the study. This needs to be a thoughtful and cost-effective process.

4.11 Baseline Data

Baseline or secondarary data have been discussed in other sections of this planning paper and thus it will be covered only briefly here.

The types of baseline data which would be suitable to gather in conducting a follow-up study of (treatment and comparison group) student consequences include the following types of data and data sources:

- Language proficiency test data on the four skills of reading, writing, speaking and listening comprehension (for both L1 and L2);
- Language dominance assessment data;

- Achievement test data;
- Affective measures of attitude, self-concept, cultural identification;
- School-level measures of success;
- Observations of student behavior in classroom and environmental settings;
- School records and documents.

For conducting follow-up studies of teachers, baseline measures might include the following:

- Language proficiency (for both L1 and L2);
- Certification data;
- Grades obtained in teacher education programs;
- Observations of the classes that the teachers teach;
- School and other records;
- Retrospectional self-report data.

When conducting follow-up studies of the extent of program implementation and institutionalization, baseline data could reasonably include the following:

- Descriptions of long term goals;
- Descriptions of annual objectives;
- Descriptions of project operating characteristics and features;
- Indicators of implementation levels;
- Funding level and pattern of funding from sources;
- Records and observations of representative classrooms;
- Records of curricula and materials used, and ratings of their emphases or quality;
- Site observations and ethnographic interviews;
- Indepth case histories.

Multiple measures (and data sources) of the processes being assessed are considered more desirable than relying on a single, "best" indicator of these processes. Multiple measures are likely to be complementary and helpful in converging on a sounder estimate of processes, as well as long term effects of program participation, than any single measure could. Also, given the highly diverse nature of bilingual programs, using multiple measures are more apt to detect long term unanticipated consequences. Such approaches also provide stronger evidence for ruling out alternative explanations for trends in data, and thus go a long way toward strengthening the internal validity of a study.

4.12 The Role of Tertiary Data

In follow-up research, tertiary data has a real and necessary role. This role is to make baseline data used more meaningful. In a sense, a stronger baseline is thereby obtained. It also is extremely helpful in clarifying trends in data which represent unanticipated consequences. Thus tertiary data can be quite useful in providing a basis for ruling out alternative explanations for the patterns of obtained data. One might also say that the importance of tertiary data increases in direct proportion to the time span elapsing between the conclusion of the original study and the onset of the follow-up data collection period. The greater the number and types of intervening events which might affect impact measures, the harder it will be to attribute such impacts to earlier program participation experiences. In perhaps the same way, the more non-equivalent that treatment and comparison groups are (due to antecedent, pre-existing differences) the greater the need for ancillary, tertiary data, to help in ruling out alternative explanations for patterns of data which potentially indicate that no effects exist from there having been a bilingual program.

In general, tertiary data may consist of: (a) unobtrusive measurements with relatively little observer intervention (e.g., content analyses of newspaper clippings), (b) retrospective/historical reviews of documents such as parent advisory council minutes and school board or tribal council meeting summaries, (c) reviews of official records (e.g., statistical analyses of achievement test scores by grade and targeted school within the school system), and (d) ethnographic interviews with key informants, a rich, insightful data source. If done with some sense of what the investigator is seeking, (a process of forming exploratory hypotheses, then hypotheses grounded in contextual factors and circumstances potentially affecting the follow-up data base) then tertiary data can have real, confirmatory and correlational "payoff."

4.13 Ethical Considerations

These also need to be addressed in designing follow-up studies. The privacy of the individual should be respected. In keeping with federal guidelines, and those developed by the American Psychological Association and similar professional organizations, certain studies should clearly point out to the individual that his/her voluntary participation is sought. No stigma or threats should be implied to potential respondents so that they feel jeopardized in some way by not participating. A honest representation of the study scope, and its potential benefits to individuals like themselves if they participate (such as leading to improvements in bilingual programs and their operations), should be made clear to those contacted.

If feasible, some incentives for participation might be offered if biases will not be introduced by such an approach. One such incentive which was used effectively by the NLS 72 Study was to regularly send participants a newsletter containing readable summaries of NLS reports and their implications for high school seniors who then were obtaining post-secondary education.

Thus permission should be obtained for individuals participating in the study. If original data already has been coded so that individuals are unable to be identified with the information they initially supplied, then some protection to the individual and/or the program already may exist. However, consent to use ethnographic films, and videotape recordings still should be sought. Consent forms have been designed for this purpose. (See Rudes, Goldsamt and Cervenka, 1980 for a discussion of such issues with respect to ethnographic data collection.) Similar consent should be sought from individuals for participating in the follow-up data collection.

Even with minimizing the respondent burden, being forthright as to the study purposes, etc., some proportion of individuals may choose not to participate. Since this may represent a form of non-random attrition, it would be wise to gather minimal background characteristics of these individuals, so that any potential effects on altering the follow-up score patterns can be understood and taken into account.

5.0 STATISTICAL AND ANALYTIC ISSUES IN CONDUCTING FOLLOW-UP RESEARCH

It is beyond the scope of this planning paper to offer detailed expositions of why certain techniques are sensible to use in analyzing follow-up data, and why others should be avoided, since they either violate key mathematical assumptions and/or introduce spurious findings. Such issues have been discussed at greater length in Morell (1979), the chapter by Reichardt in Cook and Campbell (1979) and a host of other sources. What can be given here are indications of some of the issues that pervade longitudinal, follow-up analyses. (A number of these issues actually have been introduced earlier in this planning paper because they effect other aspects of follow-up studies.)

5.1 Change scores

These appear to be intuitively useful in determining the amount of change over time, averaged over all individuals, then used in statistical significance tests. However, a number of violations of classical test theory occur when raw gain (time two minus one) or similar within-individual gain scores are used. If treatment and comparison groups are found to be equivalent (especially when randomized group assignment is used), then other approaches to assessing gain or standing of a group at a later time can be used. Prestest measures can become covariates in an analysis of covariance test. Multiple covariates also can be used to adjust posttest scores for earlier levels of standing on language dominance, achievement, socio-economic status and other variables. Such adjustments also will strengthen the power of the statistical tests for detecting a treatment effect, if one truly exists. In general, multiple regression approaches seem more advisable than gain score estimation, since, among other reasons, prediction can be based on much more information than merely the time one (baseline) variable.

5.2 Regression Efforts

However differences between the treatment and (one or more) comparison groups are apt to exist, especially when randomized group assignment or any approach which builds on program features to create such a control group (such as merely delaying the onset of entrance into the program by a semester) cannot be used. Nonrandomized (or more properly, non-equivalent) comparison groups introduce regression artifacts, in which the more extreme group of scores, upon follow-up measurement, have a tendency to change in the direction of the mean of the population from which the score was drawn. Differential regression introduces what Morell (1979) termed "pseudo effects." As he stated, "no method of statistical correction can remove all elements of regression artifact. Short of the random allocation of subjects to conditions there is no perfectly acceptable solution to the problem." However, Reichardt in Cook and Campbell (1979) has indicated various relatively useful approaches to the problem.

One such approach is to conduct multiple analyses and attempt to bracket estimates of important parameters. Multiple analyses would seek to rule out differential attrition threats, pre-existing differences, treatment-by-group confounding influences, etc. such that true treatment effects could be assessed separately from these alternative explanations.

5.3 Analysis of Covariance

For similar reasons, analysis of covariance may be inappropriate for follow-up analyses if the amount of corrective adjustment (posttest adjusted for pretest relationship or for extraneous factors) varies in unknown amounts with different non-equivalent groups. For some groups the covariance adjustment may underadjust, for other groups it may overadjust. More than one writer in the statistical literature has concluded that analysis of covariance is an overly-sensitive technique, albeit an appealing one.

5.4 Size of Effects

The issue of statistical significance versus practical significance needs clarification. The investigator needs to determine how much of a programmatically related effect appears to exist (through using statistical significance and amount of explained variation tests). However, one also needs to decide if that amount of program effect should become the basis for altering a bilingual program's directions, perhaps recommending changes in federal funding regulations, etc. (Such an event occurred in one of the writers' experience; an average of .057 of a score unit occurred in the attitude change indices averaged over 26 variables on a cohort of 2,586 first enlistees in the Navy. Because of the large sample size, this was statistically significant, but hardly one which merited changes in Naval personnel policies.)

It is also possible that there exists a "gray" area of borderline statistical significance. Certain variables, representing dimensions with a great deal of program and policy interest may lack statistical sigificance in demonstrating a trend over time. For example, perhaps the use of small numbers of students, due to the numbers of targeted students in such bilingual programs, might weaken the statistical power of the study for detecting statistical significance. However, there may still be some merits to understanding the nature of changes involved and attempting to put those changes in some form of perspective by the use of tertiary and other data sources.

Here the role of multiple analyses can be quite useful to strengthen the study analyses in other ways than by relying on statistical significance tests *per se*. (Some writers have even concluded that they have no place in educational or other evaluation areas because of the largely non-statistical needs of program evaluation.)

5.5 Trend Analyses and Time Series Analysis

While not overlooking the differential attrition effects and pre-existing group differences that may pervade the between-group comparisons which form the basic test of program effects and benefits, certain straight-forward presentations of the data can nonetheless be made, from which one might then proceed to higher order, although cautiously done, analyses.

That is, cell means or medians can be computed and plotted across the various data collection points in time. Such a plot would contain both treatment and comparison groups, so that trends in the pattern of group results can be visually inspected, and manually drawn regression lines used to gauge the relative linearity and parallelism of each set of group data points. The trend lines can offer an idea of their slopes, intercepts with the vertical axis, potential discontinuities, and intersections of one regression line with another (e.g., if the group high at one data point became much lower at a lower data point, and conversely so for another group). In the same way, the data patterns of those later leaving the program (and follow-up design) can be compared with those in the same site or design group remaining in the study. Differential attrition effects on data patterns can be estimated in this way.

Similarly, frequency distributions of qualitative or quantitative variables, percentages, and modes can be examined on a data point-by-data point basis, to visually inspect the amount of shift, the type of shift which occurred, and in what direction it occurred. Table shells, multiple bar graphs of patterns of group data points over time, and three-way cross-tabulations of bivariate relationships also can be assessed. The latter analysis would determine whether or not the relationship held for each data point and if it became stronger or weaker. These and other approaches offer ways of directly dealing with each wave of data points for a set of groups and also provide an understanding of shifts in data points, as well as how closely one group is related to the other in its average (or median) score on a given occasion.

Assuming that more than one follow-up occasion exists (i.e., the initial study plus at least one follow-up data point), then repeated measures analysis of variance, trend analysis and multiple group time series techniques may be applicable. However, as Labouvie (in Bentler, Lettieri and Austin, 1976) pointed out, certain mathematical assumptions need to be met, i.e., the variance-covariance matrix of variables across occasions needs to be homogeneous. Since, as stated earlier in this planning paper, the wider the time span between measurement occasions, the less likely that impact score changes can be attributed to program effects, then the empirical bivariate correlations of adjacent data points are apt to decrease with the size of the time span. This makes the corresponding F-ratios (to detect a statistically significant trend across occasions) become postively biased. That is, more significant trends would be found than are actually present; the investigator needs to be alert to this fact when interpreting the findings.

Trend analysis can be used with raw scores and group means. It can also apply to proportions of a group reporting a response on one measurement occasion (i.e., having inservice training made the teacher feel more confident in teaching children of limited English speaking proficiency) relative to the same question asked on other occasions (the follow-up study). The point at which the trend shifted, became optimal or declined appreciably can be detected and interpreted.

In general, time series analysis addresses the inferential evaluation question, "Does the exit from bilingual education (or graduation from school) represent a statistically significant shift, or is it a chance shift in a series following its normal course?" The comparison group shift in its own series of data points is used as a basis for determining if one group's shift differs significantly from the other. Using the comparison group, therefore, minimizes the possibility of inferring an intervention effect when there is none, or conversely, failing to identify an actual intervention effect.

Fox et al. (1976) was a very enthusiastic proponent of time series analysis. A statement he made indicates very well the benefits that can be gained from using tertiary data to explain and interpret data from those participating in bilingual education programs. To begin with, he advocated performing time series analysis, as well as preparing plots of several projects' scores across time, and on the same plot, placing indicators of key project activity onset, contextual factor occurrences, and other major "intervention" points. This forms a display of when key events began and when shifts might be attributable to each key event.

In general, Fox et al. stated: "Time series design and analysis can be an interpreter's dream: rich in evidence, provocative in perspective and ambiguous or suggestive enough for creative interpretation and real dialogue."

In his evaluation of Teacher Corps training graduates Fox and his colleagues made use of a continuous, interactive team process that periodically interpreted meaningful fluctuations of survey results and related these fluctuations to participants' expectations, actions, responses and assumptions being captured by event descriptions. In this way he may have strongly compensated for certain deficiencies in his evaluation design. "Only" 18 data points were used in the time series analysis. Other authorities advocating this method recommend 50 data points. Either level of data collection is probably impractical in a school sytem setting and highly disruptive of program operations. (Students' time on instructional task, a highly meaningful variable in bilingual evaluation, could be seriously affected by an overly extensive data collection effort.) Fox *et al*. also conceded that no preliminary (pre event) data base could serve as a baseline for the follow-up data points. Similarly, a large time lag existed between the end of the event and the beginning of follow-up surveys.

Thus time series designs are useful in evaluating the effects of a law or reform in a *post hoc* manner (e.g., the effects on institutionalization before and after the 1978 strengthening of the Bilingual Education Act as to districts having parent advisory councils and parent participation). Such designs can also be used for assessing the effect of an intervention or interventions on a single unit when many observations are feasible. Thus longitudinal tracking of: (a) an intact set of students from a small, yet highly innovative bilingual instructional program or (b) teachers who are bilingually certified to teach Native American languages, would represent a suitable and policy-relevant application of time series analysis. However, it appears to be the case that large-scale national evaluations are more the rule than the exception. In reality, small-scale focused (and complementary) studies may be more meaningful and "fit" the highly diverse field of bilingual education. Thus, the issue arises of whether or not time series studies, with many data points on relatively few individuals, are the desired perspective for policy planners. However, in general form, time series techniques may be plausible as a variation of trend analysis with its interpretative features.

5.6 Factor or Cluster Analysis

In addition to those techniques described above, other approaches can be justifiably applied to follow-up data, depending on the needs of the investigator.

If changes in structural relationships are of greater interest than quantitative changes, then factor or cluster analytic methods could be applied. In these ways relationships among multiple sets of responses within and across an occasion could be investigated, even when different sets of dependent measures are employed at different occasions. The goal would be to determine if certain variables or dimensions are less related to others on certain measurement occasions (e.g., the first follow-up) than on other occasions (e.g., the exit from the bilingual program). Thus, the pattern of factor loadings, the factor structure of which variables load on which dimensions across occasions, and indices of factorial structure similarity, such as coefficients of congruence, can be used for these purposes.

Both factor and cluster analytic methods actually represent families of mathematical algorithms, each having differing forms of logic for obtaining a set of factor dimensions, clusters of linkage structures, etc. The investigator, therefore, would be well-advised to use several approaches within each family of techniques, to assess which specific technique best explains the variation in data.

5.7 Path Analysis

This also can be useful for the same purpose of exploring the strength of certain relationships across variables and occasions, and whether definite shifts occur in those patterns. Cross-lagged correlations also can be used for similar purposes. However, neither of these can prove causality. Instead, path analysis provides a way of estimating the magnitudes of the direct and indirect influence of certain variables on other variables according to a causal ordering assumed valid by the researcher. The model itself needs to come from such sources as other studies, some consensus of bilingual local personnel, discussions with OBEMLA, or hypothesized model features which play a role in program implementation and operations (e.g., the model being currently used in the Classroom Instructional Component of the ESEA Title VII Bilingual Program study). The presumed causal relationships among the variables in the model can be tested, while the bivariate correlation coefficients can be decomposed into precise direct and indirect effects. There are various mathematical problems in applying the technique such as high degrees of correlation among conceptually related variables (multicollinearity). However, each of these issues can be checked to determine the wiseness in applying the model, given that it meets the analytic needs of the researcher.

5.8 Nonparametric Statistics

These also deserve to "have their day in court" and may be suitable for dealing with follow-up data patterns. Such techniques would be turned to when parametric techniques fall short of meeting the normality assumptions underlying the data distribution. Certain parametric tests may not be that much greater in statistical power than their nonparametric counterparts, and should not be turned to simply for the fact that they may use more of the information found in interval scales.

Nonparametric statistics were effectively used by Potter, a member of Development Associates' staff on the Experimental Schools Program study of the Edgewood Independent School District (Davis and Jones et al., 1977). The Wilcoxon Signed Ranks test and the Mann-Whitney U test were used to detect trends across small numbers of subsets of seven semesters, and treatment-groups differences. Classroom interactions, emotional climate, teacher behaviors, and teacher emphases related to programmatic emphases on communication skills were gathered via use of the Modified OSCAR classroom observation instrument. Plots and table shells were also prepared to gain an understanding of the nature of trends.

There is a definite place for nonparametric statistics, especially in bilingual education with its small numbers of targeted grades and students being served in some settings. Such techniques should not be directly rejected out of hand.

5.9 Final Thoughts on Statistical Techniques

This section has attempted to identify issues, point out pitfalls, and indicate what statistical techniques are relevant to bilingual education follow-up studies, and generally to indicate how problems might be avoided or solved. However, this section has not attempted to be a methodology and statistical analysis textbook. It has not dealt with such relevant issues as: probability or purposive sampling, questionnaire construction, psychometric issues such as score equating across test forms or levels, the use and analysis of ethnographic data (see Planning Paper 6 by Rudes et al., 1980 for such discussions), and, among other topics, pointing out what computer software already exist for conducting trend or time series analyses.

Probably what is most relevant to consider in choosing a statistical technique is that it is exactly that -- a technique and nothing more. Its value depends on its empirical validity and what numbers it has supplied. It does not provide interpretations of the practical significance and policy planning implications of findings. It cannot deal with alternative explanations of the findings if these explanations are not already supplied to it in some mathematical or quantitative form. It cannot even deal fully with findings amassed from other analyses, and reconcile both sets, e.g., jointly mesh quantitative and qualitative findings, or combine separate analyses of trends from tertiary data and follow-up data. Finally, what technique is used is the decision of the investigator and depends on what best fits the study needs and is feasible.

Multiple analyses of the same data, with the use of tertiary data when possible, may help to provide a stronger sense of what should be concluded from the data trends, and thereby provide a greater credibility and confidence in the findings to policymakers and local/state/federal program planners. In the highly diverse setting of bilingual education, such multiple analyses using alternative, complementary techniques, probably are a necessity.

6.0 SELECTION OF FOLLOW-UP VARIABLES: MEASUREMENT ISSUES

6.1 Initial Considerations

The selection of meaningful follow-up variables and sensitive measures is perhaps one of the most critical, as well as creative, tasks in planning a follow-up study. Generally, the investigator will have little or no control over baseline measurements and must use what is given. It was advocated that tertiary data be collected which bears on baseline measurements and which might possibly eluciate findings or unexpected consequences, e.g., thus having information on student characteristics may provide for reclassifying a sample population and thereby conducting more effective and insightful analysis.

It is also unlikely that the investigator can closely match up baseline and follow-up measures, e.g., the subjects taking the baseline measures might have "outgrown" the test. They may no longer be used to sensitively detect gains. "Ceiling" effects may result in score patterns. However, in follow-up studies there is no compelling need to always strive to match up instruments carefully, as in a pretest-posttest paradigm within the same study. Possibly just as fruitful, the investigator can postulate relationships (form hypotheses) between one kind of baseline measurement and another kind of follow-up measurement. The latter would appear to be the most feasible and productive type of general procedure for conducting follow-up studies in bilingual education. Again, we would urge tertiary data be collected on follow-up measurements because they are likely to be of use in elucidating findings and accounting for unanticipated consequences. Plausible alternative explanations for test results need to be considered. Such explanation may be derived from consideration of test and measure reliability, differential measurement errors, contextual differences and similar factors.

Thus, the crucial measurement issues in follow-up studies relate to (a) the selection of relevant follow-up variables and measures, (b) the decision about how to expand the "knowledge context" of these variables by assembling easily gathered tertiary data, and (c) and most importantly, posing the appro-priate research question, one which is based on clearly elaborated theory and a body of accurate empirical knowledge.

6.2 Consequences for Students

In examining the long term consequences of bilingual education for students, one is basically interested in two things: the fluctuation of treatment effects over the long term (resulting in either attentuation or "snowballing", i.e., increasing effects), and other consequences which may be due to bilingual education schooling. The evaluation design term "treatment" as applied to a bilingual education program can generally be said to consist of adding another language and culture to English monolingual education. Thus, the key variables involved in bilingual education are (a) L1 proficiency in the four language skills (speaking, listening comprehension, reading and writing); (b) L2 (English) proficiency; (c) C1 enculturation (or socialization); and (d) C2 enculturation. In follow-up studies their trends would be examined for evidence of having either attentuating or snowballing patterns.

Major consequences of bilingual education of interest to those doing follow-up study consist of academic achievement and other types of school attainment. Others, which relate to human development in general, would consist of effects on students' personality (and related to this, their attitude, motivation, self-esteem and value orientation), and their career and life choices.

Promising measures of proficiency in English and other languages appear to be on the horizon. A number of existing tests could be used with the lower grades, e.g., Boston Placement Evaluation Battery (1979) in English, Spanish and Chinese. A number of proficiency tests are likely soon to be developed based on the South West Regional Laboratory's (SWRL) Language Skills Framework (1979). For the upper grades and high school levels, there are a number of excellent tests of English proficiency, the Comprehensive English Language Test (1970), the Michigan Test of English Proficiency (1961-66) and the Test of English as a Foreign Language or TOEFL (1964-73) developed by Educational Testing Service.

To measure enculturation* a set of instruments would have to be devised, e.g., attitude scales, questionnaires, and behavioral rating scales from which C1, C2 or bicultural indices could be devised (Cervenka, 1967).

Tests of academic achievement or performance can be selected on the basis of review in A Guide to Assessment Instruments for Limited English Speaking Students (1978) and Assessment Instruments in Bilingual Education (1978).

Measures of personality factors (attitude, motivation, self-esteem and value orientation) and career and life choices can be adapted from instruments given in the two publications cited above, and also from those given in Buros' Mental Measurements Yearbook (1978).

6.3 Professional Development of Teachers

In examining the professional development of bilingual teachers from a long term perspective, follow-up studies could focus on the:

- Acquisition and maintenance of bilingual language proficiency;
- The proficiency levels required to teach in a particular language at a given grade level;
- Teacher instructional competence;
- Teacher knowledge of bilingual education theory and practices (and educational theory and practice in general);
- Attitudes toward L1 and L2, bilingual education, education, the teaching profession, and the local community; and finally,
- Teacher commitment to bilingual education as a profession.

A number of test or assessment procedures currently exist and would be suitable for assessing teachers' proficiency in a variety of languages. Teachers' instructional competence could be assessed by direct observation and by supervisor assessment. From existing sources, a comprehensive examination of teachers' knowledge of bilingual education theory and practice could be assembled and perfected. Several standardized measures of teacher attitude may also be used in a follow-up study of teachers: the Minnesota Teacher Attitude Inventory (1951) and the Bilingual Teacher Attitude Inventory (Cervenka *et al*., 1977). Other more narrowly-focused attitude scales would need to be devised to fully examine teacher attitudes and commitment to bilingual education.

---

*Enculturation can generally be defined as the way an individual learns or internalizes his/her culture(s), or as the process by which a human being adapts to his/her culture (or cultures) and learns to fulfill functions of role, status, etc.

6.4 Program Implementation and Institutionalization

Follow-up study in this areas would use the bilingual program as the primary unit of analysis. A program focus would apply various implementation measures (Fullan and Pomfret, 1977; Berman and McLaughin, 1978) to the latter phases of implementation and the onset of institutionalization within the local school system. The key implementation variables would deal with features of instruction and instructional model, while secondary variables would consist of: staff development, parent and community involvement, curriculum and material resources and program management. An additional, relevant variable would be the capacity-building activities carried out by the school system and program to facilitate the process of institutionalization. Measurement of these variables in the follow-up phase could be accomplished principally by adapting those being used in the "Evaluation of the Classroom Instruction Component of the ESEA Title VII Program" study.

7.0 LARGE SCALE POST HOC FOLLOW-UP STUDIES

The feasibility and worthiness of conducting large scale post hoc follow-up studies of bilingual students, teachers or programs requires careful scrutiny. This is because in bilingual education programs for language minority students, relevant baseline measurements may be quite diverse, limited in quantity or quality, or be difficult to obtain. The cooperation and active assistance of local school districts and programs seems essential in our view for conducting follow-up studies.

Much of the effort in conducting such follow-up studies would involve careful and lengthy identification of representative subjects with adequate baseline data, both for treatment and comparison groups, and subsequently tracking such individuals. Language minority students in bilingual programs can probably be expected to be even harder to track and locate because of their greater mobility.

Post hoc studies can stand on their own merits as a legitimate and, perhaps even preferred, study design (Strasser and Denniston, 1978). A well-developed literature exists on the use of available data for social research purposes and such data tend to form a major part of post hoc studies (Cook, 1972; Hyman, 1972). Thus it would appear that "all is not lost," from a research point of view, if follow-up research or evaluation plans cannot be built into a bilingual program at the start.

Morell (1979) points out that many occasions exist when studies are not planned at the onset of a program, or even where study procedures are not "in place" until a program is in progress or has actually terminated. The general problems of post hoc studies are well known and are particularly salient in the present context (see Sections 1.3, 4 and 5 in this report). However, Morell (1979) believes the occurrence of these problems does not preclude the still valuable use of post hoc follow-up designs: "If information is needed, if post hoc follow-up is the only available method of obtaining the information, and if sufficient care is taken to deal with difficulties, post hoc follow-up should most certainly be carried out."

Nonetheless, post hoc studies still need to deal with the limited quality of available data, the type of research questions being addressed, and the study's intrinsic value for examining meaningful fluctuations in program effects.

The major issues that post hoc follow-up studies must treat include the following:

- The problem of missing data is likely to become more severe in proportion to start-up delays in follow-up data collection;
- Missing data is likely to introduce complex forms of bias, which may be difficult to understand;
- The timing of a follow-up study may miss significant program effects, in terms of either attentuation or snowballing; and
- The problem of attrition is probably the most formidable issue faced by follow-up studies, although given current methodologies and techniques for locating follow-up subjects, it would seem that these problems can be adequately resolved.

In general, current methodologies for locating follow-up subjects, which often employ creative, detective-like investigation techniques, are effective but often are somewhat costly. (McAllister, Butler and Goe, 1973; McAllister, Goe and Butler, 1973; Crider, Willits and Bealer, 1971; Willits, Crider and Bealer, 1969; Eckland, 1968.)

7.1 Follow-up Research into Long Term Consequences for Students

The authors suggest that a long term post hoc follow-up study of students enrolled in bilingual education programs beginning in the 1966-67 school year (which were funded under Title I and Title III) might be feasible. Such students should have completed at least six years of elementary bilingual schooling. They would have graduated high school either this year or last and would be 18-20 years of age. They would also be in the process of making career and life choices. A second group which might be more amenable to long term follow-up study would be the students who entered Title VII bilingual programs in 1969-70 and who would now be eleventh graders. Two sets of comparison groups could be utilized: students of the same language and cultural background who did not participate in bilingual education and an "Anglo-American" group that also did not participate in a bilingual program.

Baseline measurements for long term student follow-up should include, if at all possible, measurements or ratings of proficiency in L1 and L2 near the beginning and end of their bilingual school and measures of academic performance. A variety of additional "tertiary" data could be collected on student and program characteristics.

Follow-up measurements would consist of language proficiency measures in L1 and L2 and a variety of school measures of success. Questionnaire and interviews could be used to obtain comprehensive information on students' social and cultural background, language skills and use, degree of enculturation (socialization) into minority group and mainstream culture and personality factors (attitude, motivation, self-esteem and value orientation). Information would also be elicited from students on their present and future life and career choices.

A significant part of follow-up data gathering would consist of collecting additional information from students by means of ethnographic interviews, in order to probe some of the more subtle effects of their participation in bilingual school. As for the choice of baseline measures, a variety of valuable teritary data would also be collected, e.g., from school documents, interviews with teachers and other school personnel.

The types of research questions that would be explored in this follow-up study would essentially consist of the illustrative questions given at the end of Section 3.1.

This proposed follow-up study might perhaps prove too costly and time-consuming to justify the possible payoffs it might produce. Also, it would perhaps be an overly ambitious attempt to recapture data on the first 12 years of bilingual education in the U.S. Therefore, a logical alternative would be to focus on a few highly significant research questions, and identify appropriate baseline and follow-up data which will fruitfully address them. Such an alternative strategy could employ baseline data from students who have participated in bilingual education programs in more recent times.

7.2 Long Term Professional Development of Bilingual Teachers

To our knowledge, no effort has been made to examine the development of bilingual teachers over an extended time perspective. Yet it is suggested that currently no other area in bilingual education would benefit more from a long term perspective. Both bilingual teacher education and inservice training would greatly benefit from long term follow-up study.

Such a follow-up study could establish a baseline by relying heavily on retrospective information, as well as documentary information on education and training. Baseline data would therefore include information on teachers' undergraduate and graduate education, their service as teacher assistants, their practicum experiences, their experience in bilingual education classes, and their inservice training on the job.

Follow-up data collection by questionnaires, structured interviews, ethnographic interviews and perhaps language testing could generate information relating to language proficiency, instructional competence, educational theory and knowledge, attitudes, and professional commitment. Supervisors' assessment of instructional competence could also be obtained; as an alternative, teacher observation could be conducted.

The types of research questions that would be explored in this teacher development follow-up study are suggested at the end of Section 3.2.

Several alternative approaches appear feasible and useful for drawing representative samples of bilingual teachers and forming comparison groups. Teachers can be compared who differ in their extent of experience, ranging from teacher interns to those with five (or more) years of active classroom experience. The latter category would necessarily exclude those who have become either resource teachers or moved upward to be administrators in the school system. Teachers can be stratified by their extent of certification and/or relevant graduate training courses, the number of hours of inservice training courses and seminars, the types of such courses taken, and/or how effective recent courses have been. The comparison groups could also take into account whether or not the teachers have ongoing experience with a particular school system, or whether they recently have come to that setting with its new linguistic/community context. What subjects the bilingual classroom teacher teaches is an important factor. Their levels of bilingual competence in L2 with respect to reading, writing and speaking should be considered. Their own linguistic behaviors in such areas as code-switching, actual language use, etc. should be taken into account in the composition of the samples. Teachers with more than two languages should be considered. Lastly, the stratification variables or other sampling approaches chosen should consider the particular language being used, i.e., Spanish or non-Spanish and what non-Spanish language is used.

Thus there are no real difficulties in establishing comparison groups and in drawing representative samples of such groups. However, the overall goals of the follow-up study should form the basis for the sample selection, so that the comparison groups are meaningful ones for drawing policy-relevant conclusions.

7.3 Program Implementation and Institutionalization

This follow-up study of program implementation and institutionalization would specifically follow-up one of the Part C Studies, "The Evaluation of the Classroom Instruction Component of the ESEA Title VII Program," which is basically a study of project implementation. Thus, it could avoid some of the hazards of unplanned, post hoc follow-up. Within one to two years after the completion of this study, there will be a need to focus a follow-up study on the later phase of implementing a bilingual program, and particularly on the processes of capacity-building and institutionalization. This follow-up study could focus on the sample of 60 projects in the indepth analysis, from which comprehensive program characteristic and implementation data are being collected through site visits. Essentially, this follow-up study would take the present Part C study one step further and answer critical questions about the long term process of institutionalization, i.e., how bilingual programs finally become incorporated in local school systems, and the trends in doing so.

Follow-up inquiry could be guided by research questions such as are given at the end of Section 3.3.

8.0 FOCUSED FOLLOW-UP STUDIES OF BILINGUAL EDUCATION CONSEQUENCES FOR STUDENTS

A series of possible follow-up studies could use as baselines the data of the Significant Bilingual Instructional Features Study and its associated special studies. Follow-up data collection on the student populations involved could begin during the final phases of these studies or after they were completed. However, planning such studies should begin as soon as feasible. Each follow-up study would have different needs for follow-up measures, observational protocols and other data gathering procedures.

The authors believe strongly that the soon-to-begin bilingual instructional features studies should be supported by small, carefully designed follow-up studies of substantive issues and problems in bilingual education. Some of these follow-up studies could perhaps be independent of the antecedent studies, and would resemble replications. However, they would all have in common an extension of "time perspective." They would both contribute to knowledge, and increase the credibility and acceptability of the instructional features studies as a whole.

These focused follow-up studies could include:

- Continued Research on Instructional Features
- Language Acquisition Follow-up
- Language Use Follow-up
- Cognitive Development Follow-up
- Case Study Follow-up of Personality Factors
- Case Study Follow-up of Students' Performance in All English Speaking Classrooms.

8.1 Continued Research on Instructional Features

It would be important to conduct as various types of "continued research" (Hyman and Wright, 1971), substudies of the major Significant Bilingual Instructional Features Study and perhaps at least two of its associated

Special Studies. One strong reason for doing so is that these follow-up studies would enhance the generalizability of the findings of the significant features studies.

Hyman and Wright (1971) have given a description of "continued research." "There are studies that build upon, extend, and occasionally test under varying circumstances the findings and hypotheses of previous research. For purposes of discussion here we shall call this practice 'continuity in research' and note that it subsumes what we earlier called comparative evaluations."

Lazarsfeld et al. (1948) emphasized the gains from continuity in research:

1. "Similar results corroborate earlier findings and thereby increase our confidence in them and their generalizability."
2. "Differences in results between two or more studies may be traced to differences in specific test conditions and thereby enable us to specify conditions under which relationships hold."
3. "Differences in results may lead to the discovery of new explanatory factors that clarify the findings."

Continued research may take different forms and pose different research questions, depending on the findings of the original studies. One cannot, of course, anticipate the type of continued research needed on this set of studies. However, it seems like that the following sorts of questions suitable for continued research might be legitimately raised:

(a) Can causal relationships between features and consequences be determined for other language-cultural groups? (not included in those studies)

(b) Can causal relationships between features and consequences be determined for similar, larger groups of students? (a form of replication issue)

8.2 Language Acquisition Follow-up

Special Study 1, Language Learning in Bilingual Instruction, seeks to determine what instructional practices foster school-related skills in the second language, and also how student language characteristics interact with bilingual instructional practices to affect acquisition of language skills. The special study is likely to cover a two-year time span. A follow-up study in that domain could use the following types of baseline data: (a) L1 and L2 proficiency (particularly English literacy), (b) language use data, (c) instructional practices data and (d) parent questionnaire data. This follow-up study could be carried out extending that study's time perspective by two to four years and, using similar follow-up measures, determine the persistence of consequences (i.e., level of language skills).

Perhaps the most important knowledge that would be produced would relate to the trend in the rate of L2 language acquisition, as a function of L1 proficiency. If strong relationships are detected in follow-up studies, there are profound implications for the educational productivity of different bilingual education approaches. Currently, interest in these relationships has been heightened by the studies of Finnish children in Sweden by Skutnabb-Kangas (1976) and Toukomaa (1977) and by Cummins' (1979) review of cognitive research on bilinguals. One could also examine the relative rates of both L1 and L2 acquisition as a function of different instructional and student characteristics.

8.3 Language Use Follow-Up

This follow-up study would focus on language used by students in actual communication, i.e., on sociolinguistic competency (Rodriguez-Brown, *et al*., 1976, 1979). Research has shown that standardized tests do not capture the language use of linguistic minority children. When these children are interviewed or observed in contexts outside of the test situations, an abundance of language use can frequently be observed. Language use data in English and perhaps Spanish could be obtained by a variety of means; by conducting ethnographic interviews (Spradley, 1979) of individual children (or even pairs or small groups of children); by obtaining such samples via wireless transmitters attached to children, and by collecting samples of their writing. Language samples would be analyzed to determine their linguistic content and the "triggers" for various instances of language use (including code-switching) and its consequences (Berney, 1972). Language samples could be linked to simultaneous observational records. Situations could be selected to concentrate on children's interactions with their peers in the classroom, on the playground, etc.

This follow-up study could utilize as a baseline that data generated by Special Study 1 and the major descriptive study, i.e., primarily L1 and L2 proficiency and language use data. A great deal of relevant tertiary data could also be brought to bear on this secondary or baseline data.

Thus, a follow-up study which focuses on language use could provide extremely useful information on student-exit, placement and educational follow-up in mainstream classrooms.

8.4 Cognitive Development Follow-Up

This study would be a follow-up of a part of Special Study 1 and perhaps aspects of the major descriptive study. One object of this follow-up study might be to explore relationships among a panel of cognitive measures and language proficiency measures. The measures could include Piagetian variables (seriation, transitivity, etc.), special measures derived from the data generated by the major significant instructional features study, psychological variables and convergence/divergence patterns of thinking. Measures could also be prepared or adapted for assessing some of the highly positive perceptual and cognitive characteristics noted among balanced bilingual children by Leopold (1939-49), Peal and Lambert (1962), and Diebold (1968), e.g., realistic perception, and non-language linked problem solving.

Special Study 1 and perhaps aspects of the major descriptive study could provide the following types of baseline data relating to cognitive development: (a) L1 and L2 proficiency, (b) cognitive aspects of language development (thinking, remembering and problem-solving, etc.) and (c) language use and cognitive skills in context.

8.5 Case Study Follow-Up of Personality Factors

This study would be a follow-up of Special Study 2, Language Attitudes of Parents, Students and Teachers in Bilingual Classrooms, and possibly also the major descriptive study.

The object of this follow-up study would be to produce case studies of a sample of children, and thereby attempt to document some of the non-quantitative and subtler effects of bilingual instruction. In particular, these would consist of personality factors or effects on attitude and motivation correlates of personality. Being a non-quantitative study, this approach would be particularly suitable to studying those language and cultural groups for which adequate assessment instruments are unavailable.

This approach would focus on producing a descriptive, largely non-interpretive account of individual bilingual children: of how they reveal themselves to an interviewer through speech and other communicative behavior, of how they appear to an observer in the classroom and elsewhere, and of how other significant individuals in the student's life (teacher, teacher aide, parent) observe and report on such students. To some extent, case studies might also rely on achievement test data and project or school records. Individual cognitive and effective assessments may also be used in describing students. Though the case study follow-up focuses on individual children, it would be important that it have secondary foci on the child's teacher and parents.

The baseline data which could be provided by Special Study 2 would consist of: (a) parents, teachers and students' attitudes toward language, culture and bilingual instruction, (b) measures of instructional features and (c) measures of classroom behavior. The major descriptive study is likely to produce a massive amount of both qualitative and quantitative data relating to such other personality factors as socialization, motivation, self-esteem and value-orientation.

8.6 Case Study Follow-Up of Students' Performance in All English Speaking Classrooms

This case study follow-up could focus on those students included in the Significant Bilingual Instructional Features Study and Special Studies No. 1 (Language Learning) and No. 2 (Language Attitudes). Students would be tracked individually until such time as when they are placed in all-English speaking or mainstream classrooms.

The overriding research question to be posed by this follow-up study would be: what school-related skills and behaviors do bilingual students most need to function effectively at various time periods in mainstream classes? The study could focus more intensively on either attitude, or language acquisition, or relationships between instructional features and consequences demonstrated in the all-English-speaking classroom, depending on which previous study students had participated in. More specifically, the research questions that could be addressed by such a study would include: What factors are the best predictors of academic and social performance in mainstream classrooms for students from bilingual classrooms? What factors should be considered in the placement of bilingual students into mainstream classrooms? What types of follow-up, educational treatments, if any, are needed by bilingual students who are placed in mainstream classrooms?

Case study methods would be primarily used in examining individual student performance in all-English speaking classrooms, including classroom observation and ethnographic interviewing of students and their new teachers. To a lesser extent, educational measurements corresponding to the baseline measurements could also be employed, along with other performance measures.

9. CONCLUSION

There is a definite place for follow-up studies in investigating the long-term, enduring and even unanticipated effects of bilingual schooling. However, these investigations will need to be based on a recognition of their value to educational audiences, relative merits of research issues, attention to the problem of attrition, proper selection of dependent measures, and the merits of planned follow-up studies. It is acknowledged that post hoc follow-up research also has merit, if that form is the more realistic to conduct. Follow-up studies also will benefit a great deal by: (a) using comprehensive and sound baseline measures, from which trends in educational effects can be appraised, and (b) annexing tertiary data as available to the follow-up data base so that trends in observed effects can be interpreted, explored in rich detail, and finally plausibly explained.

Bilingual education, as a highly diverse (perhaps the most diverse educational innovation that ever has been mounted as a federal funding initiative), could benefit a great deal from follow-up studies. If for nothing else, shorter-term effects attributable to bilingual program participation may be more difficult to detect than in other programs, given the wide diversity of students, programs, available assessment instruments and local contexts. Thus, the preliminary planning of future, long-term research should be considered, so that the benefits and educational consequences of significant features present in bilingual settings will not be underestimated and overlooked.

REFERENCES

American Institutes for Research. *Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Program, Vol. III: Year Two Impact Data, Educational Process, and In-Depth Analyses*. Palo Alto, CA, 1977.

American Institutes for Research. *Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Program: Overview of Study and Findings*. Palo Alto, CA, 1978.

Bean, J. P. *Path Analysis: The Development of a Suitable Methodology for the Study of Student Attrition*. Paper presentation at the AERA Convention, 1979.

Berman, P. and McLaughlin M. *Federal Programs Supporting Educational Change, Vol. III: Implementing and Sustaining Innovations*. R-1589/8-HEW. Santa Monica, CA: Rand, 1978.

Berney, T. D. *The Effects of Language Choice on the Task Success of Bilingual Children*. PhD dissertation, Yeshiva University, 1972.

Boston Public Schools. *Placement Evaluation Battery (PEB): A Test of All Skills -- Language Dominance and Proficiency*, 1979.

Broedling, L., Wiskoff, N. M., and Herbert, J. *Longitudinal Studies of Attitude Change: Issues and Methods* (NPRDC Special Report 75-9). San Diego, CA: Navy Personnel Research and Development Center, 1975.

Bruck, M. and Shultz, J. *An Ethnographic Analysis of the Language Use Patterns of Bilingually Schooled Children: Working Papers on Bilingualism*. No. 13, (ERIC ED 140 664), 1977.

Buros, Oscar K. (ed.). *Mental Measurements Yearbook*. 8th Ed. Highland Park, NJ: Gryphon Press, 1978.

Cazden, C. and John, V. P. Learning in American Indian Children. In: *Styles of Learning Among American Indians: An Outline for Research*. Rosslyn, VA: Center for Applied Linguistics, 1968.

Cazden, C., John, V. P., and Hymes, D. (eds.). *Functions of Language in the Classroom*. New York: Teachers College Press, 1972.

Cervenka, E. J. *Inventory of Socialization for Bilingual Children*. Austin, Texas: Child Development Evaluation and Research Center, The University of Texas at Austin, 1968.

Cervenka, E. J., Cervenka, S. B., and Orta, A. *Bilingual Teacher Attitude Inventory*. New York: New York City Board of Education, 1977.

Cohen, A. D., Fathman, A., and Merino B. The Redwood City Bilingual Education Project: 1971-74. In *Working Papers on Bilingualism*, No. 8. Toronto: Ontario Institute for Studies in Education, 1975.

Cohen, A. D. and Laosa, L. M. Second Language Instruction: Some Research Considerations. Curriculum Studies. 1976, 8(2): 149-165.

Collins, E. F. and Bailey, J. P. Jr. Survey Procedures to Maximize Participation: The National Longitudinal Study of the High School Class of 1972 (First Follow-Up). Paper presentation at the American Educational Research Association Annual Meeting, 1975.

Cook, T.D. Secondary Evaluation, Paper presented at the annual meeting of the American Psychological Association, 1977.

Cook, T. D. and Campbell, D. T. Quasi-experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally, 1979.

Crider, D. M., Willits, F. K., and Fealer, R. C. Tracking Respondents in Longitudinal Surveys. Public Opinion Quarterly. 1971, Vol. 35.

Cummins, J. Linguistic Interdependence and Educational Development of Bilingual Children. Review of Educational Research. 1979, 49, 2, pp. 222-251.

Diebold, A. R. Jr. The Consequences of Early Bilingualism in Cognitive Development and Personality Formation. In Studies of Personality: An Interdisciplinary Appraisal. (Edited by E. Norbeck et. al.) New York: Holt, Reinhart and Winston, 1968.

Droege, R. C. Effectiveness of Follow-Up Techniques in Large-Scale Longitudinal Research. Developmental Psychology. 1971, Vol. 5, No. 1, 27-31.

Eckland. B. K. Retrieving Mobile Cases in Longitudinal Surveys. In Public Opinion Quarterly. Vol. 32, 1968.

Educational Testing Service. Test of English for Foreign Language Speakers (TOEFL). 1967.

English Language Institute. Michigan Test of English Language Proficiency. Ann Arbor, University of Michigan, 1978.

Fox, G. T. Jr., Romberg, T. A., and Anglin, L. W. Reflections Upon the Use of a Time Series Design and Analysis for Program Evaluation. ERIC ED 138 650, 1976.

Fullan, M. and Pomfret A. Research on Curriculum and Instruction Implementation. Review of Educational Research. 1977, Vol. 47 (No. 1), 335-397.

Gezi, Kal. Bilingual-Bicultural Education: A Review of Relevant Research. California Journal of Educational Research. 1975, Vol. XXV, No. 5, 223-239.

Hilton, T. L. and Patrick, C. Cross-sectional Versus Longitudinal Data: An Empirical Comparison of Mean Differences in Academic Growth. Journal of Educational Measurement. 1970, 7, 1970, 15-24.

Hyman, H. H. Secondary Analysis of Sample Surveys. New York: Wiley, 1972.

Morell, J. Follow-Up Research as an Evaluation Strategy: Theory and Methodologies. In Handbook of Vocational Education Evaluation. Ed. by Abramson, T., Tittle, C. K., and Cohen, L. Beverly Hills, CA: Sage Publications, 1979.

National Center for Education Statistics. The Condition of Education 1979 - Part 2: NCES Programs and Plans. Washington, D.C.: Author, 1979.

Otal, C., Goldsamt, M. R., Wilson, L. and Hopstock, P. with Copping, B. and Toledo, F. Housing and Demographic Characteristics of the Adams Morgan - Mt. Pleasant Planning Area and Socioeconomic Characteristics of its Latino Community. Development Associates. Submitted to the Office on Latino Affairs and Department of Housing and Community Development of the District of Columbia, 1980 (Volume I: Key Study Findings; Volume II: Technical Supplement).

Palmer, L. A. and Harris, D. P. Comprehensive English Language Test (CELT). New York: Webster/McGraw-Hill Co., 1970.

Panos, R. R. and Rice, J. E. Experimental Analysis of Different Mail Follow-up Techniques Within a National Sample of Hard-Core Nonrespondents. Proceedings of the Annual Convention of the American Psychological Association, 1967. (Summary)

Part C Coordinating Committee. Proposed Research Plan for Bilingual Education. Education Division, DHEW, July 1979.

Peal, E. and Lambert, W. E. The Relation of Bilingualism to Intelligence, Psychological Monographs: General and Applied. 1962, Vol. LXXVI, No. 27, 1962.

Pletcher, B. P., Locks, N. A., Reynolds, D. F., and Sisson, B. G. A Guide to Assessment Instruments for Limited English Speaking Students. New York: Santillana Publishing Co., Inc., 1978.

Rodriguez-Brown, F. V. et al. Language Interaction in a Bilingual Classroom: An Observational Study. Paper presented at 1976 AERA meeting, San Francisco, 1976.

Rodriguez-Brown, F. V. The Effect of Language Used for Early Reading Instruction: A Bilingual Perspective. Paper presented at the 1979 AERA meeting, San Francisco, 1979.

Rodriguez-Brown, F. V., Bruck, M. Walcer, C. S., and Shultz, J. Language Interaction in a Bilingual Classroom: An Observational Study. Paper presented at the 1976 AERA meeting, San Francisco, 1976.

Rudes, B. A., Goldsamt, M. R., and Cervenka, E. J. Issues in Applying Ethnography to Bilingual Education: Planning Paper 6. Arlington, VA: Development Associates, 1980. Prepared for the National Institute of Education under subcontract to Abt Associates, Inc.

Saldate, M. and Mishra, S. Cognitive Benefits of a Spanish-English Bilingual Education Program. Paper presented at NABE Conference, San Juan, P.R., 1978. Duplicated.

Hyman, H. H. and Wright, C. R. Evaluating Social Action Programs in F. G. Caro (ed.) Readings in Evaluation Research. New York: Russell Sage Foundation, 1971.

Jones, E. & Davis, P. B. with Theimer, W. C., Cervantes, R. A., Goldsamt, M. R., Toscano, H. X., and Bernal, E. Volume I - Final Summary Report on the Experimental Schools Project, Edgewood Independent School District. (San Antonio, Texas). Development Associates. Submitted to the National Institute of Education, 1977.

Kim, J. and Curry, J. The Treatment of Missing Data in Multivariate Analysis. In Alwin, D.F. (ed.) Survey Design and Analysis: Current Issues. Beverly Hills, CA: Sage, 1978.

Labouvie, E. Longitudinal Designs. Chapter 4 in Bentler, P. M., Lettieri, D. J., and Austin, G A. (eds.). Data Analysis Strategies and Designs for Substance Abuse Research. Rockville, MD: National Institute on Drug Abuse, 1976.

Lazarsfeld, P. F. et. al. The People's Choice. 2nd edition, New York: Columbia University Press, 1948.

Leopold, W. F. Speech Development of a Bilingual Child: A Linguist's Record. (Northwestern University Studies, Humanity Series No. 6, 11, 18, 19) Evanston, IL: Northwestern University Press, 4 vols., 1939-49.

Locks, N. A., Pletcher, B. A., and Reynolds, Dorothy F. Language Assessment Instruments for Limited English-Speaking Students: A Needs Analysis. Washington, D.C.: National Institute of Education, 1978.

Marcus, A. C. et. al. An Analytical Review of Longitudinal and Related Studies as They Apply to the Educational Process. In Methodological Foundations for the Study of Several Effects, Vol. 3. Washington, D.C.: National Center for Educational Statistics, 1972.

Marshall, G. Methods for Minimizing Attrition in Field Studies. April, 1976. ED 128 412.

McAllister, R. J., Coe, S. J. and Butler, E. W. Tracking Respondents in Longitudinal Surveys: Some Preliminary Considerations. In Public Opinion Quarterly. 1973, Vol. 37, No. 3.

McCord, J. A Thirty-Year Follow-Up of Treatment Effects. In Sechrest, L. et al. Evaluation Studies: Annual Review. Beverly Hills, CA: Sage Publications, 1979.

Messick, S. "The Critical Problem in the Evaluation of Instruction: Assessing Possible, Not Just Intended, Outcomes. In Wittock, M. C. and Wiley, D. E. (eds.), The Evaluation of Instruction: Issues and Problems. New York: Halt, 1970.

Minnesota Teacher Attitude Inventory. New York: The Psychological Corporation, 1951.

Schaie, K. W. A General Model for the Study of Developmental Problems, Psychological Bulletin. 1965, 64, 92-107.

Scriven, M. Maximizing the Power of Casual Investigations: The Modus Operandi Method. In Glass, G. V. (ed.) Evaluation Studies: Annual Review. Beverly Hills, CA: Sage Publications, 1975.

Silverman, R. J., Noa, J. K., and Russell, R. H. Oral Language Tests for Bilingual Students: An Evaluation of Language Dominance and Proficiency Instruments. Portland, Oregon: Center for Bilingual Education, Northwest Regional Educational Laboratory, 1976.

Skutnabb-Kangas, T., and Toukomaa, P. Teaching Migrant Children's Mother Tongue and Learning the Language of the Host Country in the Context of the Socio-Cultural Situation of the Migrant Family. Helsinki: Finnish National Commission for UNESCO, 1976.

Sobel, S. B. Throwing the Baby Out with the Bathwater: The Hazards of Follow-up Research. In American Psychologist. Vol, 33, 1979.

Spradley, J. P. The Ethnographic Interview. Chicago, IL: Holt, Rinehart, and Winston, 1979.

St. Pierre, R. G. Planning Longitudinal Field Studies: Considerations in Determining Sample Size. Evaluation Review, 1980, Vol. 4, No. 3, 405-415.

Strasser, S. and Denniston, O. L. Pre and Post Planned Evaluation: Which is Preferable? Evaluation and Program Planning. 1978, Vol. 1, No. 3. 1978.

SWRL. Language Skills Framework: Vol. 1 Oral Language; Vol. 2 Reading; and Vol. 3 Writing. 1979.

SWRL. Language Assessment Manual: Resources for Developing a Student Placement System for Bilingual Programs. 1980.

Tucker, G. R. Bilingual Education - Some Perplexing Observations. Educational Evaluation and Policy Analysis. 1(5), Sept. - Oct. 1979.

Toukomaa, P. and Skutnabb-Kangas, T. The Intensive Teaching of the Mother Tongue to Migrant Children of Pre-School Age. Research Reports 26. Tampere, Finland: Department of Sociology and Social Psychology, University of Tampere, 1977.

Willits, F. K., Crider, D. M., and Bealer, R. C. A Design and Assessment of Techniques for Locating Respondents in Longitudinal Sociological Studies. Washington, D.C.: Center for Epidemiologic Studies, DHEW, 1969.